GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Der Einfluss des Abstands zwischen Initialschätzung und Ratschlag auf die Ratschlagsnutzung

Dissertation

zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades

"Doctor rerum naturalium"

der Georg-August-Universität Göttingen

im Promotionsprogramm ‚Grundprogramm Biologie'

der Georg-August University School of Science (GAUSS)

vorgelegt von

Anne-Fernandine Rakotoarisoa

aus Greifswald

Göttingen, 2012

Mitglieder des Betreuungsausschuss:

Prof. Dr. Stefan Schulz-Hardt
(Abteilung für Wirtschafts- und Sozialpsychologie, Georg-Elias-Müller-Institut für Psychologie)

Prof. Dr. Michael Waldmann
(Abteilung für Kognitionswissenschaft und Entscheidungspsychologie, Georg-Elias-Müller-Institut für Psychologie)

Dr. Thomas Schultze
(Abteilung für Wirtschafts- und Sozialpsychologie, Georg-Elias-Müller-Institut für Psychologie)

Mitglieder der Prüfungskommission:

Referent: Prof. Dr. Stefan Schulz-Hardt

Korreferent: Prof. Dr. Michael Waldmann

Weitere Mitglieder der Prüfungskommission:

Prof. Dr. Margarete Boos
(Abteilung für Sozial- und Kommunikationspsychologie, Georg-Elias-Müller-Institut für Psychologie)

Prof. Dr. Uta Lass
(Abteilung für Experimentelle Psychologie, Georg-Elias-Müller-Institut für Psychologie)

Prof. Dr. Uwe Mattler
(Abteilung für Experimentelle Psychologie, Georg-Elias-Müller-Institut für Psychologie)

Prof. Dr. Hannes Rakoczy
(Abteilung für Biologische Entwicklungspsychologie, Georg-Elias-Müller-Institut für Psychologie)

Tag der mündlichen Prüfung: 22.10.2012

Danksagung

An erster Stelle gilt mein Dank meinem Doktorvater Stefan Schulz-Hardt. Sein brillanter analytischer Sachverstand sowie seine Klarheit und Präzision beim Erdenken und Umsetzen experimenteller Fragestellungen haben mich durch die unterschiedlichen Phasen meiner Promotion geleitet und stets zutiefst beeindruckt. Des Weiteren möchte ich Michael Waldmann für die Zweitbegutachtung meiner Doktorarbeit danken. Ich danke außerdem Margarete Boos, Uta Lass, Uwe Mattler und Hannes Rakoczy dafür, dass sie sich als Mitglieder der Prüfungskommission für meine Dissertation zur Verfügung gestellt haben.

Zu tiefem Dank bin ich meinem Kollegen und Freund Thomas Schultze verpflichtet, der als Anleiter meiner Doktorarbeit fungierte und der einen nicht unerheblichen Beitrag zur Entstehung dieser Arbeit geleistet hat. Ich danke Thomas für seine stets vorhandene Hilfsbereitschaft, das Teilen seiner genialen Gedanken sowie die unterstützende und wertschätzende Atmosphäre, in der unsere Zusammenarbeit stattfand.

Ebenso bedanken möchte ich mich bei all den Personen, ohne deren Hilfe und Engagement die Durchführung der Experimente, welche ich in meiner Dissertation berichte, nicht möglich gewesen wäre. So gebührt mein Dank den (überwiegend ehemaligen) studentischen Hilfskräften Anette Opielka, Anna Gast, Annika Giersiepen, Maresa Wendtland und Dario Özkent sowie natürlich den zahlreichen Versuchspersonen.

Besonderer Dank gilt meinen Freunden, die den so wichtigen Ausgleich zum wissenschaftlichen Arbeiten darstellen und die mich stets durch ihre soziale Unterstützung und Zuwendung inspiriert, gefördert, aber auch - in sehr positivem Sinne - abgelenkt haben. Auch danke ich meinen Eltern für ihre immerwährende Unterstützung und den Glauben an meine Fähigkeiten.

Astrid kann ich eigentlich gar nicht genug danken, versuche es aber dennoch.

Inhaltsverzeichnis

1. Einleitung 1

2. Theoretischer und empirischer Hintergrund 4

2.1 Das Judge-Advisor-System 5

2.2 Aufgabentypen im Judge-Advisor-System und Bestimmung der Ratschlagsgewichtung 7

2.3 Einflussfaktoren auf die Ratschlagsgewichtung 12

3. Ableitung der Fragestellung 16

3.1 Einfluss der Ratschlagsdistanz auf die Ratschlagsgewichtung 17

3.2 Soziale Validierung als Aspekt der Ratschlagsnutzung 24

4. Experiment 1 29

4.1 Zielsetzung und Hypothesen 29

4.2 Methode 30

4.2.1. Stichprobe und Design des Experimentes 30

4.2.2 Ablauf 30

4.3 Ergebnisse und Diskussion 34

4.3.1 Analyse möglicher Störvariablen 34

4.3.2 Ratschlagsgewichtung 35

4.3.3 Urteilssicherheit 42

4.3.4 Explorative Datenanalyse: Schätzungsakkuratheit der Judges 44

5. Experiment 2 49

5.1 Zielsetzung und Hypothesen 49

5.2 Methode 50

5.2.1. Stichprobe und Design 50

5.2.2 Ablauf 50

5.3 Ergebnisse und Diskussion 51

5.3.1 Analyse möglicher Störvariablen 51

5.3.2 Ratschlagsgewichtung 52

5.3.3 Urteilssicherheit 55

5.3.4 Explorative Datenanalyse: Schätzungsakkuratheit der Judges 57

6. Experiment 3 61

6.1 Zielsetzung und Hypothesen 61
6.2 Methode 64
6.2.1 Stichprobe und Design 64
6.2.2 Ablauf 65
6.3 Ergebnisse und Diskussion 66
6.3.1 Analyse möglicher Störvariablen 66
6.3.2 Ratschlagsgewichtung 66
6.3.3 Urteilssicherheit 71
6.3.4 Explorative Datenanalyse: Schätzungsakkuratheit der Judges 73
7. Experiment 4 77
7.1 Zielsetzung und Hypothesen 77
7.2 Methode 80
7.2.1 Stichprobe und Design 80
7.2.2 Ablauf 81
7.3 Ergebnisse und Diskussion 82
7.3.1 Analyse möglicher Störvariablen 82
7.3.2 Ratschlagsgewichtung 83
7.3.3 Urteilssicherheit 87
7.3.4 Explorative Datenanalyse: Schätzungsakkuratheit der Judges 91
8. Übergreifende Diskussion 95
8.1 Einfluss der Ratschlagsdistanz auf die Ratschlagsgewichtung 96
8.2 Einfluss der Ratschlagsdistanz auf die Urteilssicherheit 101
8.3 Adjustierung versus soziale Validierung 102
8.4 Schätzungsakkuratheit der Judges 103
8.5 Beschränkungen und Ausblick auf die weitere Forschung 106
8.6 Schlussbemerkung 109
9. Literatur 110
Anhang 116

1. Einleitung

Beratungssituationen sind ein wesentlicher Bestandteil der menschlichen Entscheidungsfindung. Angesichts einer zu treffenden Entscheidung – ob nun mit weitreichenden Konsequenzen wie etwa Geldanlagen oder die Karriereplanung oder aber schlicht den Tagesablauf betreffend – holen sich Menschen gern und häufig den Ratschlag anderer ein; insbesondere dann, wenn die Ratgeber[1] hinsichtlich des betreffenden Bereichs über eine gewisse Expertise verfügen. Ratschläge fungieren als zusätzliche Informationen und schließen Wissenslücken, wodurch die Informationsbasis, auf Grund derer bestimmte Entscheidungen getroffen werden, so sorgfältig und vollständig wie möglich gestaltet wird und anschließende Entscheidungen mit größerer Sicherheit getroffen werden können. Diese Vorgehensweise ist sinnvoll, denn die Nutzung von Ratschlägen trägt häufig zu einer Verbesserung der Qualität eines Urteils oder einer Entscheidung bei (Sniezek, Schrah & Dalal, 2004; Yaniv & Milyavsky, 2007; Harvey & Fischer, 1997). Dass das Ziel der Entscheidungsoptimierung ein häufig verfolgtes Gut darstellt, zeigt sich nicht zuletzt darin, dass Menschen sehr oft dazu bereit sind, viel Geld für das Einholen einer zusätzlichen Meinung, Information oder Einschätzung auszugeben.

Das Urteilen und Entscheiden im Rahmen sozialer Kontexte, insbesondere mit Hilfe von Ratschlägen, Meinungen oder Informationen anderer, weist also eine hohe Relevanz im menschlichen Alltagsleben auf. Allerdings erfreut sich diese Thematik erst eines recht jungen Forschungsinteresses. Lange Zeit beruhte Forschung zum Urteilen und Entscheiden zumeist entweder auf individuellen Entscheidungsparadigmen oder auf Modellen zu Gruppenentscheidungen. Beratungssituationen sind in der Regel aber weder vollkommen individuell, noch verteilt sich die Verantwortlichkeit für eine zu treffende Entscheidung oder ein zu treffendes Urteil gleichmäßig auf alle am Prozess beteiligten Personen. Vielmehr ist die Verantwortlichkeit in Berater-Entscheider-Kontexten sehr einseitig verteilt und liegt (zumeist) bei einem Individuum allein. Der

[1] Zur Vereinfachung der Lesbarkeit wird im Folgenden das generische Maskulinum verwendet. Sofern nicht explizit anders ausgeführt, sind damit jedoch stets auch weibliche Personen gemeint.

Entscheidungs- oder Urteilprozess erfolgt aber dennoch meist nicht in sozialer Isolation. Das Judge-Adisor-System (Sniezek & Buckley, 1995) überträgt diese Konstellation der Entscheidungsfindung auf laborexperimentelle Untersuchungen. Es bildet Situationen ab, in denen ein Entscheidungsträger Empfehlungen oder Informationen eines Ratgebers einfordert oder auch ungefragt erhält, um sie zu einem Gesamturteil zu integrieren (Sniezek, 1999), und ermöglicht dadurch, die unterschiedlichsten Variablen im Zusammenhang mit der Ratschlagsverarbeitung zu untersuchen.

Der überwiegende Bereich der Forschung mit Anwendung des Judge-Advisor-Paradigmas fokussiert auf das Ausmaß, in dem erhaltene Ratschläge von einem Entscheidungsträger berücksichtigt werden. Generell legt die Empirie dabei den Schluss nahe, dass Ratschläge in geringerem Ausmaß berücksichtigt werden, als es ihre objektive Qualität nahelegen würde. Dieses so genannte *Advice Discounting* (z.B. Yaniv, 2004a; für einen Überblick Bonaccio & Dalal, 2006) wird von den Autoren allerdings nicht ausschließlich für die Bezeichnung einer suboptimalen bzw. Unternutzung von Ratschlägen verwendet, sondern gelegentlich auch, um jedwede Abweichung von einer 100%-Gewichtung von Ratschlägen zu Gunsten der eigenen Einschätzung zu beschreiben (z.B. Yaniv, 2004b; Yaniv & Kleinberger, 2000). Von besonderer Bedeutung im Sinne möglicher Beeinträchtigungen für die Entscheidungs- und Urteilsqualität sind allerdings natürlich vor allem die Fälle, in denen Ratschläge gemessen an ihrer Güte in nur unzureichendem Umfang berücksichtigt werden. Eine begriffliche Unterscheidung beider Prozesse ist daher sinnvoll und wird im Rahmen dieser Arbeit vorgenommen. Um auszudrücken, dass ein Ratschlag ungeachtet seiner Güte nicht vollständig übernommen wird, wird der Bergriff *Advice Discounting* bzw. *Diskontierung* verwendet. Wenn ein Ratschlag stattdessen in geringerem Maße gewichtet wird, als dies gemessen an seiner objektiven Qualität adäquat wäre, verwende ich den Begriff *Unternutzung*.

Im Zusammenhang mit der Ratschlagsgewichtung allgemein wurden bereits diverse Variablen überprüft und identifiziert, die beeinflussen, inwieweit Ratschlägen Folge geleistet wird. Ein wiederkehrend gefundenes und intuitiv einleuchtendes Ergebnis ist dabei, dass Ratschläge stärker berücksichtigt werden, wenn derjenige, der

sie erteilt, eine hohe bzw. in Relation zum Entscheidungsträger höhere aufgabenbezogene Kompetenz oder gar Expertise besitzt (z.B. Harvey & Fischer, 1997), aber auch zahlreiche andere Faktoren, wie etwa die wahrgenommene Ähnlichkeit zum Ratgeber (Gino, Shang & Croson, 2009) oder aber die subjektive Sicherheit bezüglich seiner Ratschläge (Sniezek & Buckley, 1995; Sniezek & Van Swol, 2001) spielen eine Rolle für das grundlegenden Ausmaß der Ratschlagsgewichtung (für einen ausführlichen Überblick Bonaccio & Dalal, 2006).

Einen weiteren wesentlichen Einflussfaktor auf die Ratschlagsgewichtung stellt der Ratschlag selbst dar; konkret, sein Abstand zu einem anfänglichen Urteil. Die besondere Notwendigkeit der Betrachtung dieses Ratschlagsmerkmals ergibt sich aus zwei Gründen: Zum einen liegt diese Information – die Existenz einer Anfangsmeinung/-annahme vorausgesetzt – im Gegensatz zu solchen über den Ratgeber und/oder die objektive Qualität seiner Ratschläge immer vor. Menschen haben dadurch stets die Gelegenheit, erhaltene Ratschläge hinsichtlich der Ähnlichkeit oder Unterschiedlichkeit zu ihrer Anfangseinschätzung zu bewerten. Zum anderen ist bereits aus der *Attitude Change*-Forschung bekannt, dass sich die Unterschiedlichkeit einer anfänglichen Meinung und einer zusätzlich erhaltenen von Menschen wahrgenommen wird und sich systematisch auf das Ausmaß der Einstellungsänderung auswirkt (z.B. Bochner & Insko, 1966; Hovland, Harvey & Sherif, 1957). Die Schlussfolgerung liegt nahe, dass dies auch im Kontext von Entscheider-Berater-Situationen der Fall sein dürfte. Zur Generierung von Annahmen über die Beziehung zwischen Ratschlagsentfernung und –gewichtung wurde daher die Forschung im Rahmen des Judge-Advisor-Systems mit der Attitude Change-Forschung in Verbindung gebracht (Yaniv, 2004b). Bisherige Befunde von Minson, Liberman und Ross (2011) sowie Yaniv (2004b) zeigen, dass mit zunehmendem Abstand zwischen Ratschlag und der Anfangsschätzung eines Entscheidungsträgers die Gewichtung der erhaltenen Ratschläge abnimmt.

Da außer einer Untersuchung von Yaniv (2004b) bis dato keine weitere Studie vorliegt, die im Rahmen des klassischen Judge-Advisor-Systems systematisch die Effekte der Entfernung von Anfangsurteilen und Ratschlägen auf die Ratschlagsgewichtung untersucht, soll das Ziel meiner Dissertation darin liegen, eine umfassendere Erkennt-

nis dieses Zusammenhangs zu erlangen. In einer Reihe von vier Experimenten untersuche ich, welche Art des Zusammenhangs zwischen der Ratschlagsentfernung und -gewichtung durch Menschen, die für ein finales Urteil verantwortlich sind, besteht. Ich überprüfe weiterhin, ob die gefundenen Zusammenhänge für verschiedene Aufgabentypen, Kompetenzlevel der Ratgeber und Arten der Distanzmanipulation variieren oder stabil bleiben. Zusätzlich soll parallel zur Ratschlagsgewichtung auch die Urteilssicherheit erfasst und ebenfalls auf ihre Abhängigkeit von der Ratschlagsdistanz, Ratgeberkompetenz oder Distanzmanipulation hin überprüft werden. Durch die Erhebung der Urteilssicherheit sollen Erkenntnisse darüber gewonnen werden, ob bestimmte Entfernungen von Ratschlägen mit Prozessen der sozialen Validierung einhergehen. Dies geschieht mit dem Ziel, soziale Validierung durch Ratschläge neben der reinen Ratschlagsgewichtung als einen weiteren möglichen Aspekt der Ratschlagsnutzung zu identifizieren.

2. Theoretischer und empirischer Hintergrund

Wenn man bedenkt, wie allgegenwärtig Beratungssituationen in unserem täglichen Leben sind, ist es doch ein wenig überraschend, dass die Untersuchung der Beeinflussung von Urteilen und Entscheidungen durch Ratschläge noch ein so recht junges Forschungsfeld darstellt. Das dazugehörige Paradigma, das Judge-Advisor-System, so wie es auch in meiner Dissertation Anwendung findet, wurde erst Mitte der Neunziger Jahre von Sniezek und Buckley (1995) in die Forschungsliteratur eingeführt. Der überwiegende Teil der Forschung unter Anwendung dieses Paradigmas befasst sich seitdem mit dem Ausmaß der Berücksichtigung erhaltener Ratschläge bzw. der Gewichtung von Ratschlägen (z.B. Harvey & Fischer, 1997; Yaniv, 2004b; Yaniv & Kleinberger, 2000) und damit verbundenen Einflussfaktoren seitens des Entscheiders und des Ratgebers. Weiterhin wird sich mit den Charakteristika der Kommunikations- bzw. Interaktionsmöglichkeiten und –gegebenheiten befasst, die sich auf die Annahme (oder Ablehnung) von Ratschlägen auswirken sollen (z.B. Feng & MacGeorge, 2006; Harvey & Fischer, 1997; Schrah, Dalal & Sniezek, 2006; Van Swol & Sniezek, 2005).

Mit den Faktoren, die das Befolgen eines Ratschlags beeinflussen, werde ich mich auch im Rahmen meiner Doktorarbeit befassen. Allerdings stehen dabei nicht nur Charakteristika des Ratgebers im Fokus, sondern auch und hauptsächlich Eigenschaften des Ratschlags selbst. Mit Verweis auf die Attitude Change-Literatur wird bisher davon ausgegangen, dass Ratschläge umso weniger gewichtet werden, je weiter entfernt sie von einer initialen Entscheidung oder einem initialen Urteil liegen. Die Kenntnis über den Zusammenhang zwischen Ratschlagsentfernung und -gewichtung gilt es im Folgenden einer kritischen Prüfung zu unterziehen. Zuvor stelle ich dafür das zur Untersuchung von Beratungssituationen gebräuchliche Forschungsparadigma (Kapitel 2.1) sowie die darin verwendeten unterschiedlichen Aufgabentypen und die gängigen Maße zur Bestimmung der Ratschlagsgewichtung (Kapitel 2.2) vor. Im Anschluss daran umreiße ich die wesentlichen Befunde zur Ratschlagsgewichtung und deren Einflussfaktoren (Kapitel 2.3). Danach befasse ich mich detailliert mit dem Verhältnis von Ratschlagsentfernung und -gewichtung und erläutere, warum ich der Auffassung bin, dass die bisherigen Annahmen bezüglich dieses Zusammenhangs erweitert bzw. ergänzt werden müssen (Kapitel 3.1). Darauf aufbauend werde ich darlegen, warum ich weiterhin der Ansicht bin, dass die Ratschlagsgewichtung nicht in jedem Fall genügt, um das volle Ausmaß der tatsächlichen Ratschlagsnutzung zu erfassen. In diesem Zusammenhang werde ich in Kapitel 3.2 Überlegungen zu Ratschlägen als Quelle sozialer Validierung darlegen und diese ebenfalls in Bezug zur Ratschlagsentfernung setzen. Die innerhalb dieser beiden Kapitel entwickelten Forschungsfragen werden mit Hilfe von vier Experimenten (Kapitel 4 – 7) überprüft und beantwortet sowie abschließend diskutiert (Kapitel 8).

2.1 Das Judge-Advisor-System

Die Anforderungen an ein Paradigma, welches Beratungssituationen laborexperimentell untersuchbar machen soll, sind relativ klar: Es soll eine –zumindest schematische– Abbildung eines natürlichen Entscheidungsprozesses liefern, der beinhaltet, dass die Entscheidungsgewalt sowie in Folge die Verantwortlichkeit für getroffene Entscheidungen einerseits überwiegend bei einer Einzelperson liegen, Entscheidungen aber dennoch andererseits selten in sozialer Isolation getroffen werden bzw. das Ein-

holen separater Meinungen zumeist auch solchen Individualentscheidungen vorausgeht. Diese Forderungen werden im Rahmen des Judge-Advisor-Systems (JAS; Sniezek & Buckley, 1995) erfüllt und dadurch Beratungsvorgänge im Labor generiert und untersuchungsfähig gemacht.

Der Begriff Judge wird dabei für die Teilnehmer, die eine Entscheidung oder ein Urteil treffen sollen, verwendet. Im Laufe eines Versuchs erhalten diese Judges dann -aufgefordert oder unaufgefordert- Ratschläge von anderen Versuchsteilnehmern, die der Rolle des Advisors, also der des Ratgebers, zugeordnet worden sind. Die Ratschlagsangabe kann auch non-interaktiv erfolgen – wobei es sich um den Normalfall der Durchführung handelt. In diesem Falle werden Ratschläge dann meistens computerbasiert dargeboten. Dabei werden sie untersuchungsabhängig künstlich generiert oder aber echte Ratschläge früherer Probanden via PC dargeboten. Die Zuteilung zur Rolle des Judges oder des Advisors erfolgt in den allermeisten Fällen zufällig.

Das JAS besteht in der Regel aus drei Phasen: Zunächst gibt der Judge ein erstes Urteil (= Initialurteil) oder eine erste Entscheidung ab. Im Anschluss daran erhält er einen aufgabenbezogenen Ratschlag. In einem dritten Schritt erhält er abschließend die Möglichkeit ein endgültiges und möglicherweise revidiertes Urteil (= Finalurteil) abzugeben. Dieser dreistufige Vorgang läuft dabei zumeist nicht nur ein einziges Mal ab, sondern wird für eine beliebige bzw. nötige Anzahl von Versuchsdurchgängen wiederholt. In manchen Untersuchungen (z.B. Yaniv, 2004b) geben die Judges zunächst alle Initialurteile auf einmal ab und erhalten dann im zweiten Schritt alle Ratschläge, um schließlich in der dritten Phase ihre gesamten Finalurteile abzugeben. In anderen Untersuchungen hingegen läuft jeder einzelne Urteilsprozess sofort dreistufig (Initialurteil-Ratschlag-Finalurteil) ab. Die tatsächliche Ausgestaltung des Paradigmas unterliegt in der empirischen Praxis zahlreichen Modifikationen und hängt natürlich nicht zuletzt von den untersuchten Parametern ab (Sniezek & Buckley, 1995). Selbst die Durchführung als dreistufiges Paradigma ist nicht obligatorisch. So verzichten manche Untersuchungen beispielsweise auf die Abgabe eines Initialurteils oder einer Initialentscheidung. Weiterhin erfolgt nach Verteilung der Rollen meist noch die vom Ziel der Untersuchung geleitete Manipulation der unabhängigen Variablen. Auch die Interakti-

on bzw. Interaktionsmöglichkeiten von Judge und Advisor, oder aber die Anzahl der Advisors, von denen ein Judge Ratschläge erhält, gestalten sich ganz unterschiedlich (als Überblick Bonaccio & Dalal, 2006).

2.2 Aufgabentypen im Judge-Advisor-System und Bestimmung der Ratschlagsgewichtung

Wie sich bereits im vorherigen Kapitel angedeutet hat, kann bei den im Rahmen des JAS verwendeten Aufgaben, ähnlich wie in der Kleingruppenforschung (z.B. Gigone & Hastie, 1997), eine Unterscheidung zwischen Urteils- und Entscheidungsaufgaben getroffen werden. Bei Entscheidungsaufgaben sollen Probanden eine Wahl aus mehreren qualitativ unterschiedlichen Alternativen treffen. Realisiert werden diese Fragen bzw. Aufgaben häufig in Form einer Mehrfach-Wahlaufgabe (Multiple-Choice-Frage). Ein erhaltener Ratschlag beinhaltet dann die Empfehlung einer der nicht-gewählten bzw. nicht-präferierten Alternativen oder aber eine mit der Ausgangsentscheidung übereinstimmende Empfehlung. Unabhängig davon, ob vor der Darbietung des Ratschlags eine Initialentscheidung getroffen werden soll oder nicht, kann die Bestimmung der Ratschlagsgewichtung hier nur auf die Art erfolgen, dass überprüft wird, ob Ratschlag und Finalentscheidung übereinstimmen bzw. „matchen“ (*„matching“*; aus Sniezek & Buckley, 1995) oder nicht. Dementsprechend kann ein Ratschlag entweder nur vollständig übernommen oder vollständig diskontiert werden. Bei der Übernahme eines Ratschlags kann dann noch unterschieden werden, ob die resultierende Finalentscheidung von der möglicherweise zu Beginn zu treffenden Initialentscheidung abweicht oder aber einfach diese anfängliche Meinung, auch als Folge eines sie unterstützenden Ratschlags, beibehalten wurde. Da die Ausprägung einer Initialentscheidung wie erwähnt an sich aber keine Rolle bei der Bestimmung des Ausmaßes des matching spielt, kann es im Falle einer Ratschlagsdiskontierung bei Entscheidungsaufgaben mit mehr als zwei Alternativen Bedeutungsunterschiede geben. So kann es einerseits zu einer Diskontierung gekommen sein, weil die eigene Initialentscheidung beibehalten wurde. Alternativ ging mit der Diskontierung eventuell auch der Wechsel auf eine andere dritte bzw. beliebige Entscheidungsalternative einher, die weder der Initialentscheidung noch dem Ratschlag entspricht. Numerisch gesehen drücken sich

beide Fälle von Diskontierung im matching gleich aus, unterscheiden sich jedoch in ihrer Qualität bzw. Bedeutung. Dies sollte dann bei der Auswertung berücksichtigt werden.

Wenn die Aufgabe im JAS nicht darin besteht, zwischen vorgegebenen Alternativen zu wählen, sondern stattdessen darin, quantitative Einschätzungen oder Vorhersagen auf einer bestimmten kontinuierlichen Skala, wie etwa Jahreszahlen von historischen Ereignissen, Gewichte in Gramm oder aber Temperaturen in Grad Celsius etc. zu treffen, handelt es sich um Urteilsaufgaben. Ein erhaltener Ratschlag kann auch in solch einem Fall vollständig übernommen oder abgelehnt werden, man kann sich ihm aber auch im Rahmen der durch ihn und dem eigenen Initialurteil vorgegebenen Grenzen annähern, und man kann auch über diese Grenzen hinausgehen.

Bonaccio und Dalal (2006) kritisieren in ihrem Überblicksartikel, dass in vielen Untersuchungen zum *Judgment and Decision Making* keine Erklärungen dazu erfolgen, warum ein spezifischer Aufgabentyp verwendet wurde. Dabei hätte sich in verschiedenen Untersuchungen (z.B. Gigone & Hastie, 1993, 1997; Hinsz, 1999; zitiert nach Bonaccio & Dalal, 2006, S. 138) gezeigt, dass der Antwortmodus den Entscheidungsprozess beeinflusst. Darum möchte ich hiermit ausführen, dass ich im Rahmen meiner Dissertation Urteilsaufgaben verwende, das heißt, die Probanden meiner vier Experimente waren jeweils dazu aufgefordert, quantitative Einschätzungen auf vorgegebenen Skalen zu vorgegebenen Fragestellungen abzugeben. Dies hat sowohl methodische als auch praktische Gründe. Im Hinblick auf die Distanzeffekte, die ich zu untersuchen gedenke, wäre es auch nach reiflichen Überlegungen meiner Meinung nach nicht möglich, zur Beantwortung der Untersuchungsfrage geeignete Wahl- bzw. Entscheidungsaufgaben zu konstruieren. Einerseits können qualitative Wahlmöglichkeiten nicht oder nur sehr schwer systematisch in einem vordefinierten Ausmaß von einer Initialentscheidung abweichen, da sie eben keine quantitativen Möglichkeiten sind, und dadurch ist es andererseits auch nicht möglich variierende Distanzen als unabhängige Variablen zu manipulieren. Um Aussagen über eine graduelle und sich eventuell verändernde Ratschlagsgewichtung innerhalb von Personen treffen zu können erschien es daher unumgänglich Urteilsaufgaben zu verwenden.

Zur Quantifizierung des tatsächlichen Ausmaßes der Ratschlagsgewichtung im Rahmen von Urteilsaufgaben gibt es unterschiedliche formelbasierte Ansätze (für eine Übersicht über die verschiedenen Methoden siehe auch Bonaccio & Dalal, 2006):

Zum einen gibt es das *Advice Taking* (AT; Harvey & Fischer, 1997), bei welchem die Gewichtung eines Ratschlages als das Verhältnis der Differenz zwischen Final- und Initialschätzung und der Differenz von Ratschlag und Initialschätzung definiert ist:

$$\text{AT} = \frac{\text{Finalschätzung} - \text{Initialschätzung}}{\text{Ratschlag} - \text{Initialschätzung}}$$

Das AT-Maß bildet somit die prozentuale Gewichtung eines Ratschlags ab. Es kann positive und negative Werte annehmen. Negative Werte deuten auf eine Bewegung des Finalurteils entgegen des Ratschlags hin, treten in der Realität jedoch sehr selten auf (Gino, Shang & Croson, 2009). Werte größer 1 sind ebenfalls möglich und resultieren, wenn die Finalschätzung eines Judges in Richtung des Ratschlages tendiert, aber über diesen zusätzlich noch hinausgeht. Ein AT-Wert von 0 ergibt sich, wenn ein Ratschlag völlig unberücksichtigt bleibt; und wenn der Judge einen Ratschlag vollständig übernimmt, erreicht das AT einen Wert von 1. Resultiert ein AT von 0.5, hat der Judge den Mittelwert aus seinem Initialurteil und dem erhaltenen Ratschlag gebildet, um zu einem Finalurteil zu gelangen. Jede Art von Bewegung in Richtung eines Ratschlags (ohne diesen zu erreichen) wird durch Werte zwischen 0 und 1 angegeben.

Yaniv (2004) entwickelte mit dem *Weight of Advice* (WOA) ein ähnliches Maß zur Bestimmung der Ratschlagsgewichtung. Das WOA ergibt sich nämlich ebenfalls aus dem Verhältnis der Differenzen aus Final- und Initialschätzung sowie Ratschlag und Initialschätzung, mit dem Unterschied, dass jeweils die Beträge der Differenzen in Nenner bzw. Zähler stehen. Mathematisch ausgedrückt bedeutet dies:

$$\text{WOA} = \frac{|\text{Finalschätzung} - \text{Initialschätzung}|}{|\text{Ratschlag} - \text{Initialschätzung}|}$$

Dementsprechend kann das WOA ausschließlich positive Werte annehmen, wobei ein WOA von 0 auch hier bedeutet, dass ein erhaltener Ratschlag vollständig

ignoriert wird, während ein Wert von 1 eine vollständige Übernahme des Ratschlags als Finalschätzung widerspiegelt. Werte zwischen 0 und 1 sprechen für eine mehr oder minder stark ausgeprägte Annäherung an den dargebotenen Ratschlag, mit 0.5 als Ausdruck für die Mittelung von Initialschätzung und Ratschlag um zu einem Finalurteil zu gelangen. Diese Aussage, dass jedes WOA zwischen 0 und 1 ein gewisses Ausmaß an Ratschlagsannahme quantifiziert, ist allerdings nicht uneingeschränkt gültig, denn durch die Verwendung absoluter Differenzen kann sich bei diesem Ansatz eine Interpretation des erhaltenen Koeffizienten schwierig gestalten, wenn die Finalschätzung eines Judges außerhalb des durch Ratschlag und Initialschätzung vorgegebenen Range liegt. Wie bereits oben erwähnt, sind diese Fälle eher selten bzw. geben Judges in ca. 95% der Fälle Finalschätzungen ab, die zwischen Initialurteil und Ratschlag liegen (z.B. Gino, 2008; Gino et al., 2009), aber während beim AT-Maß negative Werte auf so einen Fall hindeuten, kann das WOA dies nicht abbilden. Jedwede Veränderung von Initial- zu Finalschätzung mutet somit als Ratschlagsgewichtung an, unabhängig davon, ob tatsächlich eine Urteilsverschiebung in Richtung des Ratschlags stattgefunden hat. Ein Anpassung der Finalschätzung über den erhaltenen Ratschlag hinaus kann wiederum allerdings auch das WOA mit Werten größer 1 abbilden.

Yaniv und Kleinberger (2000) verwendeten in ihrer Untersuchung das Gegenteil des WOAs, sprich die Gewichtung der eigenen Schätzung des Judges. Dieses *Weight of Own Estimate* (WOE) ist das Verhältnis der absoluten Differenzen von Ratschlag und Finalschätzung sowie Ratschlag und Initialschätzung:

$$\text{WOE} = \frac{|\text{Ratschlag} - \text{Finalschätzung}|}{|\text{Ratschlag} - \text{Initialschätzung}|}$$

Mit diesem Maß, welches sich außerdem aus der Differenz von 1 – WOA ergibt, kann die Gewichtung der eigenen Schätzung des Judges quantifiziert werden. Demgemäß bedeutet ein Wert von 0 in diesem Fall, dass einem Ratschlag vollständig gefolgt wurde, während ein Wert von 1 die vollständige Beibehaltung der eigenen initialen Schätzung anzeigt. Werte zwischen 0 und 1 zeigen die prozentuale Gewichtung der Initialschätzung im Hinblick auf ein getroffenes Finalurteil gegenüber dem erhaltenen Ratschlag an.

Neben diesen formelbasierten Ansätzen zur Bestimmung der Ratschlagsgewichtung kann diese bei Urteilsaufgaben auch mit Hilfe regressionsbasierter Ansätze erfolgen. Ratschläge und –falls Teil des verwendeten Paradigmas– Initialschätzungen fungieren dann als Prädiktoren für die Finalschätzung der Probanden, die wiederum das Kriteriumsmaß zur Bestimmung einer Regressionsgleichung darstellt. Auf diese Art und Weise kann dann die Gewichtung erhaltener Ratschläge sowie eventuell der Initialschätzung, aber auch anderer im Versuchsablauf manipulierter Variablen bestimmt werden. Wobei *Gewichtung* im Falle regressionsbasierter Ansätze nicht wie bei den formelbasierten Ansätzen mit einer Verschiebung des Finalurteils in Richtung des Ratschlags gleichzusetzen ist, sondern mit dem Ausmaß der gemeinsamen Kovariation von Prädiktoren (z.B. Initialschätzungen und Ratschläge) mit dem Kriterium (Finalurteil). Wenn jemand also beispielsweise so vorginge, die dargebotenen Ratschläge immer durch die Zahl 10 zu dividieren, um auf diese Weise zu seinen Finalurteilen zu gelangen, so ließen sich diese Finalschätzungen sehr gut an Hand des Ratschlags vorhersagen, weil beide Größen perfekt miteinander korrelierten. Die erhaltenen Ratschläge würden aus einem regressionsbasierten Blickwinkel somit vollständig genutzt werden, obwohl sie durchgängig diskontiert worden sind. Die Verwendung korrelativer Maße ermöglicht dieser Methode andererseits die Bestimmung der aufgeklärten Kriteriumsvarianz durch die erhaltenen Ratschläge bzw. durch die vorhandenen Prädiktoren. In diesem Zusammenhang wird auch ein besonderer Vorteil dieser Methode gegenüber den formelbasierten Ansätzen deutlich, nämlich, dass im Falle mehrerer vorhandener Ratschlagsquellen eine anteilige Gewichtung dieser am Gesamt- bzw. Finalurteil bestimmt werden kann. Dem gegenüber steht zugleich der Nachteil bzw. die Notwendigkeit, dass der Judge zu diesem Zwecke stets eine Reihe von verschiedenen Urteilen abgeben muss, damit eine Regressionsgleichung berechnet werden kann, wohingegen bei den formelbasierten Ansätzen die Gewichtung eines erhaltenen Ratschlags pro Untersuchungsdurchgang berechenbar ist. Weiterhin ergibt sich im Zusammenhang mit regressionsbasierten Ansätzen das Problem der Multikollinearität, wenn die Prädiktoren, also die Schätzungen der Judges und der/des Ratgeber/s, nicht unabhängig voneinander sind.

Nach Abwägung der mit den unterschiedlichen Methoden zur Bestimmung der Ratschlagsgewichtung bei quantitativen Aufgaben verbundenen Vor- und Nachteile werde ich das AT-Maß zur Bestimmung der Ratschlagsgewichtung verwenden. Gegenüber dem WOA birgt dies den Vorteil, dass die Ratschlagsannahme tendenziell nicht überschätzt wird und natürliche Schwankungen im Gewichtungsprozess zum einen dadurch sichtbar sind und zum anderen auch in die durchschnittliche Gewichtung mit eingehen. Dieser Umstand ermöglicht es dadurch eventuell, ein realistischeres Abbild des Gewichtungsausmaßes zu erlangen. Gegenüber der Verwendung regressionsbasierter Ansätze ermöglicht mir das AT-Maß weiterhin, auch intraindividuelle Veränderungen der Ratschlagsgewichtung über mehrere Versuchsdurchgänge hinweg zu erfassen, was sich als unumgänglich für die Untersuchung von Distanzeffekten erweist.

2.3 Einflussfaktoren auf die Ratschlagsgewichtung

Der Motivation, Entscheidungen und Urteile möglichst akkurat zu fällen, folgend holen Menschen die Ratschläge anderer Personen ein. Eine Fülle von Untersuchungen konnte die Verbesserung der Urteilsqualität durch die Kombination verschiedener Entscheidungen und Urteile bereits nachweisen (bspw. Gardner & Berry, 1995; Sniezek et al., 2004; Yaniv, 2004a). Neben der Akkuratheit getroffener Urteile und Entscheidungen richtet die Forschung zur Ratschlagsannahme ihr Augenmerk vorwiegend auf den Aspekt bzw. das Ausmaß der Ratschlagsgewichtung. In diesem Zusammenhang wurden bereits zahlreiche Faktoren den Advisor, den Judge oder die sie umgebende Situation betreffend dahingehend untersucht, wie sie die Ratschlagsannahme beeinflussen. Im laborexperimentellen Kontext zeigt sich auf der einen Seite, dass Probanden erhaltene Ratschläge nicht gänzlich ignorieren können und sie stets zu einem gewissen Prozentsatz gewichten, unabhängig davon wie unterschiedlich der Expertise- bzw. Trainingsstatus zwischen Judge und Advisor ausfällt. Dieser so genannte *„token Amount“* (zu dt. in etwa: symbolischer Betrag) beläuft sich in der Untersuchung von Harvey und Fischer (1997) auf ca. 20% bis 30% Ratschlagsgewichtung. Die Gründe für eine solche „Grundgewichtung“ könnten darin liegen, dass die Judges ein erhaltenes Hilfsangebot nur ungern ausschlagen, insbesondere im Hinblick auf zukünftige Situationen, in denen sie dann bei vorausgehender Zurückweisung der Ratschläge eine Vor-

enthaltung derselben befürchten (Sniezek & Buckley, 1995). Weiterhin erlaubt die Annahme von Ratschlägen eine gewisse Aufteilung der Verantwortung zwischen Judge und Advisor (Harvey & Fischer, 1997) – dies ist allerdings natürlich eine eher subjektive bzw. gefühlte Verantwortungsaufteilung, da, wie bereits erwähnt, im JAS per definitionem der Judge die Verantwortung für finale Urteile und Entscheidungen trägt. In diesem Zusammenhang ebenfalls bemerkenswert ist, dass auch bei besonders invaliden Ratschlagsquellen wie etwa Zufallszahlen noch immer eine positive Gewichtung gefunden wird, welche signifikant verschieden von 0 ausfällt (Schultze, Mojzisch & Schulz-Hardt, 2012). Als Erklärungsansätze wurden auch hier verschiedene Möglichkeiten in Betracht gezogen und bestätigt: Einerseits, dass Menschen die Hypothese, dass der wahre Wert in Richtung des dargebotenen Ratschlags liegt, unwillkürlich und positiv testen, woraufhin das Wissen, das dafür spricht, aktiviert wird und anschließend besser verfügbar ist, um diese Hypothese dann tatsächlich zu bestätigen (u.a. Mussweiler & Strack, 1999) . Andererseits erfolgte die Nutzung invalider Ratschläge auch intendiert, weil bei den Judges eine Fehlwahrnehmung bezüglich der Validität derartiger Ratschläge vorliegt. So gaben einige Versuchspersonen vor den Urteilsdurchgängen an, die Gewichtung einer Zufallszahl führe zu besseren Schätzungen und sollte daher in Finalurteile integriert werden. Eine solche Fehlwahrnehmung der Validität konnte zusätzlich noch verstärkt werden, wenn die Zufallszahlen in einem plausiblen Wertebereich schwankten.

Neben der anscheinend vorhandenen Grundgewichtung von Ratschlägen existiert allerdings der robuste Befund, dass Menschen quantitative Ratschläge zu einem gewissen Ausmaß auch immer *nicht* befolgen, das heißt diskontieren. In der überwiegenden Mehrheit der Untersuchungen werden Ratschläge also nur selten vollständig übernommen. Wie bereits eingangs erwähnt, besteht in diesem Zusammenhang jedoch eine gewisse definitorische Unschärfe zwischen Advice Discounting im Sinne jeglicher Zurückweisung bzw. Nicht-Gewichtung von Ratschlägen einerseits und einer *übermäßigen* Diskontierung (und somit suboptimalen Ausnutzung) andererseits.

Hinsichtlich letzterer besteht nun das Problem, dass ein Übermaß an Diskontierung vorweg erst einmal festgelegt werden muss. In Ermangelung einheitlicher Krite-

rien zur Bestimmung einer Soll-Gewichtung von Ratschlägen müssen vorerst Plausibilitätsüberlegungen zur Interpretation der Angemessenheit einer Ratschlagsgewichtung genügen. So empfiehlt sich bei gleichem Kompetenzstatus von Judge und Advisor etwa eine 50%-Soll-Gewichtung der Ratschläge bzw. die Bildung des arithmetischen Mittels aus beiden Urteilen (Soll & Larrick, 2009). Dementsprechend sollten die Ratschläge eines, gemessen am Vorwissen des Judges, kompetenteren Ratgebers stärker als 50% und im gegenteiligen Falle weniger als 50% gewichtet werden. Konkrete Werte einer Soll-Gewichtung lassen sich in diesen Fällen allerdings ohne vorherige Kenntnis des Leistungsniveaus beider Beteiligter nicht bestimmten.

Beeinträchtigungen der Entscheidungs- und Urteilsqualität sind dann zu erwarten, wenn Ratschlägen gemessen an ihrer Qualität nur ungenügend Folge geleistet wird. In der Literatur existieren verschiedene Erklärungsansätze für das Phänomen der Unternutzung von Ratschlägen. Zum einen könnte sie aus einer „Anchoring & Adjustment“-Strategie (Tversky & Kahneman, 1974) resultieren, bei der die Initialschätzung des Entscheidungsträgers als Anker fungiert, welcher nach Darbietung eines Ratschlages (in nur unzureichendem Maße) in Richtung dieser Empfehlung adjustiert wird (Lim & O’Connor, 1995). Zum anderen haben Entscheidungsträger Zugang zu ihren eigenen (internalen) Überlegungen und können somit gegenüber sich selbst begründen, wie sie zu einer bestimmten Entscheidung oder zu einem bestimmten Urteil gelangt sind. Im Hinblick auf die Überlegungen und Gedankengänge des Ratgebers ist dies jedoch nicht oder nur deutlich weniger der Fall. Auf Grund dieser Informationsasymmetrie wird das eigene Initialurteil stärker gewichtet (Yaniv, 2004a, 2004b; Yaniv & Kleinberger, 2000). Drittens ist Advice Discounting möglicherweise die Folge eines egozentrischen Bias, welcher besagt, dass Entscheider ihre eigenen Urteile und Entscheidungen bevorzugen, weil sie diese denen von anderen (also auch Ratgebern) für überlegen halten (Clement & Krueger, 2000; Krueger, 2003).

Unabhängig davon, welche Definition zu Grunde gelegt wird, wurden verschiedenste Faktoren identifiziert, die Advice Discounting beeinflussen oder konkret sogar vermindern. Ratschläge einer Expertenquelle führen so beispielsweise zu einer größeren Beeinflussung des Judges und seines Finalurteils, oder mit anderen Worten

zu einem weniger starken Advice Discounting, als wenn Judge und Advisor gleich oder aber der Advisor weniger kompetent ist als der Judge (z.B. Harvey & Fischer, 1997; Jungermann & Fischer, 2005). Ebenso förderlich für die Ratschlagsgewichtung sind ein höheres Alter bzw. eine höhere Lebenserfahrung, Bildung und Weisheit des Advisors gegenüber dem Judge (Feng & MacGeorge, 2006). Yaniv, Choshen-Hillel und Milyavksy (2011) zeigten, dass Ratschläge von Ratgebern der gleichen Altersgruppe, des gleichen Geschlechts und religiösen Hintergrunds wie der Judge stärker von letzterem gewichtet wurden als Ratschläge von Ratgebern mit anderer Gruppenzugehörigkeit. Auch bei Gino et al. (2009) zeigte sich der positive Einfluss von empfundener Ähnlichkeit hinsichtlich Alter, Geschlecht und Bildungsstand zwischen Judge und Advisor auf die Ratschlagsgewichtung. Wenn der Ratgeber sich zudem sehr zuversichtlich im Hinblick auf seine dargebotene Hilfestellung zeigt, erhöht dies ebenfalls die Ratschlagsgewichtung (Van Swol & Sniezek, 2005). Faktoren wie Ähnlichkeit hinsichtlich des sozioökonomischen Status oder aber das vom Ratgeber angegebene Vertrauen in seine Hilfestellung sind natürlich keine direkten Cues für die gute oder schlechte Qualität eines Ratschlages. Sie sind vermutlich vielmehr die Grundlage für Annahmen, die der Judge hinsichtlich der Kompetenz des Ratgebers trifft.

Neben Charakteristika der beiden am Urteils- oder Entscheidungsprozess beteiligten Personen sowie des sie umgebenden Kontexts sind auch Merkmale des Ratschlags von Bedeutung. Die Zahl der Untersuchungen, in denen man sich bis dato mit diesbezüglichen Fragestellungen befasst hat, ist jedoch recht überschaubar (siehe Bonaccio & Dalal, 2006). Zum einen kann aus ihnen geschlossen werden, dass bei Bereitstellung mehrerer Ratschläge pro Urteil diejenigen stärker gewichtet werden, die nahe beieinander und in der Nähe zum Initialurteil liegen. Einzelne Ratschläge, die der Mehrheit der anderen zu Verfügung stehenden Ratschläge widersprechen, werden dagegen weniger stark im Finalurteil berücksichtigt (Harries, Yaniv & Harvey, 2004). Wenn nur ein Ratschlag zur Verfügung gestellt wird, wird dieser in Relation zum Initialurteil gesehen. Hier geht die Forschung bislang davon aus, dass mit zunehmender Entfernung desselben zum Initialurteil auch die Diskontierung zunimmt bzw. anders formuliert, dass Ratschläge umso weniger gewichtet werden, je weiter entfernt sie vom Initialurteil liegen (Minson et al., 2011; Yaniv, 2004b, siehe auch Kapitel 3.1). Aus Er-

gebnissen zur Forschung im Themenkomplex des *Belief Updatings* lässt sich zudem ableiten, dass neben der Distanz auch Ratschlagsmerkmale wie die Komplexität und der Umfang (z.B. die Anzahl von enthaltenen Teilinformationen) der dargebotenen Informationen deren Integration in ein Endurteil beeinflussen (Hogarth & Einhorn, 1992).

Zusammenfassend lässt sich nach den vorangegangenen Schilderungen festhalten, dass die meisten Untersuchungen Merkmale des Judges und/oder des Ratgebers hinsichtlich ihres Einflusses auf die Ratschlagsgewichtung untersuchten. Nur in wenigen Studien wurden (zusätzlich) Charakteristika der Ratschläge einer näheren Betrachtung unterzogen. Die Ratschlagsentfernung ist so ein ratschlagsspezifischer Einflussfaktor. Im Rahmen meiner Dissertation soll es nun darum gehen, ihren Einfluss auf die Ratschlagsgewichtung zu untersuchen und damit bereits vorhandene empirische Erkenntnisse einer kritischen Prüfung zu unterziehen.

3. Ableitung der Fragestellung

Da die Qualität von Ratschlägen variiert, besteht für die Entscheider eine erhebliche Herausforderung darin, gute von schlechten Ratschlägen zu unterscheiden. Die Finanzkrise der erst kürzlich vergangenen Jahre war (und ist nach wie vor) eine deutliche Veranschaulichung dessen, was mit dem Geld der Menschen passieren kann, die den falschen Investitionshinweisen folgen bzw. den falschen Beratern vertrauen. Wie zuvor ausführlich dargelegt, konnte die Forschung dabei zeigen, dass sich Entscheidungsträger in großem Maße an sogenannten harten Fakten hinsichtlich eines Ratgebers oder seiner Ratschläge orientieren, wie etwa die Ratgeber-Expertise bzw. -kompetenz (Harvey & Fischer, 1997) oder aber – insbesondere bei mehrstufigen Entscheidungsprozessen – durch Feedback zu bereits getroffenen Teilentscheidungen (Yaniv & Kleinberger, 2000). In Abwesenheit solcher harten Leistungsdaten müssen sich Entscheidungsträger auf eher indirekte Indikatoren für die Ratschlagsqualität verlassen, wie etwa die Reputation eines Beraters (Yaniv & Kleinberger, 2000), sein Alter oder sein sozialer Status (Feng & MacGeorge, 2006).

Judges können zusätzlich aber auch noch Charakteristika des Ratschlags selbst nutzen. In gewissen Situationen könnte dies notwendig sein, weil relevante Informationen über den Ratgeber, wie die über seine Kompetenz, seine bereits erbrachten Leistungen oder seinen Ruf, nicht zugänglich sind. Stellen Sie sich zum Beispiel den neuen Arzt oder Bankberater vor, den Menschen das erste Mal sehen und daher noch gar kein umfassendes Bild von seiner Expertise erworben haben können. Ebenfalls könnten Charakteristika des Ratschlags selbst auch als zusätzliche Entscheidungsgrundlage dienen, und zwar weil in der Regel davon auszugehen ist, dass die Leistung eines Ratgebers im Laufe eines Beratungsprozesses variieren kann. Das heißt, auch sehr kompetente Ratgeber könnten unter Umständen einmal weniger gute Ratschläge erteilen. Demgemäß ist davon auszugehen, dass Charakteristika des Ratgebers insbesondere hilfreich sind, wenn es darum geht, die durchschnittliche Qualität der von ihm erteilten Ratschläge einzuschätzen. Charakteristika des Ratschlags selbst –wie etwa die Distanz zur eigenen Initialschätzung– liefern möglicherweise inkrementelle Informationen über jeden einzelnen erhaltenen Ratschlag und stellen deshalb eventuell ein zusätzliches Qualitätsmerkmal zur wahrgenommenen Expertise dar. Ein sehr markantes und wichtiges Merkmal ist in diesem Zusammenhang die Entfernung von Ratschlägen zu den (Ein)Schätzungen oder Annahmen der Entscheider (Yaniv, 2004b; Yaniv & Milavsky, 2007).

3.1 Einfluss der Ratschlagsdistanz auf die Ratschlagsgewichtung

Meiner Kenntnis nach gibt es bis zum heutigen Zeitpunkt nur sehr wenige Studien, die die Auswirkung der Entfernung zwischen Initialschätzung und Ratschlag auf dessen Gewichtung systematisch untersuchen (Minson et al., 2011; Yaniv, 2004b). Yaniv (2004b) argumentiert, dass das Ausmaß, zu dem ein bestimmter Ratschlag berücksichtigt bzw. gewichtet wird, mit zunehmendem Abstand zwischen diesem Ratschlag und einem Anfangsurteil einer Person abnimmt. Yaniv trifft diese Behauptung in Anlehnung an ein Experiment aus der Forschung zur Einstellungsänderung. Während Distanzeffekte auf die Ratschlagsgewichtung im Rahmen des JAS nämlich noch eher spärlich untersucht sind, wurden die Auswirkungen der Diskrepanz zwischen einer eigenen und einer zweiten Meinung auf die Nutzung der letzteren im Rahmen der Attitude

Change-Forschung bereits mit großer Intensität und Studiendichte überprüft. Insbesondere in den späten 50er- und den 60er-Jahren erfreute sich diese Thematik einer großen Beliebtheit, und verbunden mit dieser entstand eine hohe Zahl an Publikationen zu diesem Thema (z.B. Aronson, Turner & Carlsmith, 1963; Bochner & Insko, 1962; Brewer & Crano, 1968; Cohen, 1959; Fisher & Lubin, 1958; Freedman, 1964; Insko, Murashima & Saiyadain, 1966). In der von Yaniv (2004b) zitierten Studie von Bochner und Insko (1966) wurde seiner Ansicht nach gezeigt, dass das Verhältnis von der Diskrepanz zwischen einer persuasiven und einer Ausgangsmeinung und dem Umfang der Einstellungsänderung einem monoton fallenden Trend folgt.

In seiner eigenen Studie ließ Yaniv (2004b) seine Versuchsteilnehmer Jahreszahlen bestimmter historischer Ereignisse (z. B. Jahr der ersten Öffnung des Suez-Kanals) auf der Grundlage ihrer eigenen ersten Schätzung und im Anschluss erhaltener Ratschläge schätzen. In den beiden Experimenten, welche Teil seiner Untersuchung zum Distanzeffekt waren, wurden den Versuchsteilnehmern einmal reale und einmal künstlich erzeugte Ratschläge dargeboten. Diese Ratschläge wurden hinsichtlich ihrer Entfernung zum Initialurteil kategorisiert: So unterschied Yaniv zwischen „nahen“, „intermediären“ und „fernen“ Ratschlägen.

Im ersten Experiment dienten echte Schätzungen früherer Probanden als Ratschläge. Dabei ordnete ein Computer diese Ratschläge aus einem vorhandenen Pool von jeweils 120 Schätzungen entsprechend ihrer absoluten Entfernung (in Jahren) zur Initialschätzung an. Im Anschluss wählte er Schätzungen des 20. Perzentils als „nahe“ Ratschläge. Das heißt 20% der Schätzungen aus dem Pool lagen in diesem Falle zwischen der Initialschätzung des Probanden und dem letztendlich ausgewählten Ratschlag. Ratschläge des 55. Perzentils wählte der Computer als „intermediäre“ Ratschläge aus und „ferne“ Ratschläge entnahm er dem 90. Perzentil. Die auf diese Weise jeweils ausgewählten Ratschläge wichen im Mittel zu 24.1 (nahe Ratschläge), 50.1 (intermediäre Ratschläge) bzw. 93.8 Jahre (ferne Ratschläge) von den Initialschätzungen der Probanden ab. Im zweiten Experiment wurden die Ratschläge künstlich erzeugt. Dazu wurden je nach Abstandskategorie jeweils drei Konstanten von den ursprünglichen Schätzungen abgezogen oder addiert, welche auf Grundlage der mittleren abso-

luten Abweichungen des vorangegangenen Experiments gebildet wurden. Bei nahen Ratschlägen waren die Konstanten 15, 18 oder 20 Jahre, bei intermediären Ratschlägen 40, 43 oder 45 Jahre, und bei fernen Ratschlägen wurden Werte von 70, 72 oder 75 Jahren addiert oder subtrahiert. Die von Yaniv prognostizierte monoton fallende Beziehung zwischen Ratschlagsentfernung und -gewichtung konnte von ihm in den zwei aufeinanderfolgenden Experimenten empirisch nachgewiesen werden. Außerdem zeigte sich, dass bei Aufteilung der Probanden in kompetente (*„high-knowledge“*) und weniger kompetente (*„low-knowledge“*) Versuchspersonen an Hand ihrer mittleren absoluten Abweichung in der Initialphase, weniger kompetente Probanden dieses monoton fallende Gewichtungsmuster nicht so stetig zeigten. Stattdessen scheinen sie Ratschläge manchmal recht unabhängig von ihrer Abweichung zur Initialschätzung im relativ gleichen Ausmaß zu nutzen. Yaniv (2004b) zufolge geschieht diese teilweise unterschiedliche Gewichtung in Abhängigkeit der Ratschlagsdiskrepanz vermutlich auf Grund unterschiedlicher Attributionen bezüglich des Zustandekommens dieser Diskrepanz zwischen der eigenen Schätzung und der eines anderen: Während kompetentere Probanden hierfür die Ursache in einer fehlerhaften Schätzung des anderen, also des Ratgebers, sehen, attribuieren weniger kompetente Probanden eine solche Abweichung auf ihre eigene fehlerhafte Initialschätzung.

Minson et al. (2011) konnten in einer Untersuchung zu dyadischer Zusammenarbeit bei quantitativen Urteilsaufgaben zeigen, dass Probanden die Schätzungen eines Partners nicht gebührend gewichteten, und dies insbesondere dann, wenn die Schätzungen sich stark voneinander unterschieden. Allerdings entspricht die Versuchsdurchführung in dieser Studie nicht dem klassischen Ablauf des JAS. So wurden dort pro Versuchsdurchgang über vier Runden hinweg Urteile gefällt - zunächst erfolgte eine Initialschätzung, anschließend sollte eine zweite Schätzung mit Kenntnis der Schätzung des Partners abgegeben werden, gefolgt von einem gemeinschaftlichen Konsens-Urteil nach Diskussion bzw. Meinungsaustausch oder dem Austausch von Geboten und zuletzt sollte noch einmal ein individuelles Finalurteil getroffen werden. Weiterhin waren die Rollen der Dyaden-Mitglieder nicht klar in Judge und Advisor aufgeteilt, sondern beide Personen erfüllten gegenseitig beide Rollen (füreinander), und dementsprechend gab es auch keine einseitige Verantwortungsaufteilung.

Nichtsdestotrotz konnten somit also zwei sehr unterschiedliche Untersuchungen eine ähnliche Beziehung zwischen Ratschlagsdistanz und Ratschlagsgewichtung ermitteln. Dass dieser Befund aber unter Umständen gar nicht so gesichert ist, zeigt interessanterweise die Untersuchung von Bochner und Insko (1966), auf die sich Yaniv (2004b) zur Herleitung seiner Annahmen bezieht. Die Autoren fanden eine lineare Beziehung zwischen Diskrepanz und Einstellungsänderung nur bei einem Teil ihrer Probanden, nämlich dann, wenn persuasive Informationen durch eine sehr glaubwürdige Quelle erteilt wurden. Dieser lineare Trend war allerdings – ganz im Gegensatz zur Darstellung von Yaniv (2004b) – ansteigend, das heißt, dass zunehmend diskrepante Meinungen auch zu einer zunehmenden Einstellungsänderung führten. Wenn eine bestimmte Information bzw. Meinung jedoch durch eine Quelle mittlerer Glaubwürdigkeit kundgetan wurde, ergab sich eine kurvilineare Beziehung zwischen der Diskrepanz von Entscheider und Informant einerseits und der Meinungsänderung andererseits. Diese kurvilineare Beziehung war umgekehrt U-förmig und implizierte einen geringen Einfluss der Quelle bei Meinungen, die der Meinung der Versuchsteilnehmer entweder sehr ähnlich oder sehr unähnlich waren. Im Vergleich dazu hatten moderat abweichende Meinungen einen höheren Einfluss. Bei der gefundenen kurvilinearen Beziehung handelte es sich im Übrigen um die von den Autoren in ihrer Studie zur Einstellungsänderung ursprünglich vorhergesagte. Ihre Argumentation für diese Hypothese leitete sich dabei aus den Annahmen der Social Judgment Theory (Sherif & Hovland, 1961) ab, die ebendiese kurvilineare Beziehung zwischen den beiden Dimensionen vorhersagt.

Nach der Social Judgment Theorie besteht eine individuelle Haltung bzw. Meinung, die Inhalt einer persuasiven Kommunikation ist, immer aus einem Annahme- und einem Ablehnungsbereich bzw. einem *Latitude of Acceptance* und einem *Latitude of Rejection*. Fällt eine persuasive Botschaft in den Annahmebereich, wird sich dieser Botschaft mit zunehmender Diskrepanz auch in größerem Umfang angenähert. Fällt sie dahingegen in den Ablehnungsbereich, nimmt die Annäherung an sie mit wachsender Diskrepanz ab. Ein kurvilinearer Zusammenhang zwischen Diskrepanz und Meinungsänderung ist die Folge. Mit anderen Worten führt eine überzeugende Botschaft im Annahmebereich in aller Regel zu einer Meinungsänderung, was nur geringfügig bis

überhaupt nicht der Fall ist, wenn sie als in den Ablehnungsbereich fallend wahrgenommen wird. In Abhängigkeit der Diskrepanz zum eigenen Standpunkt wird ein persuasives Argument vom Individuum dahingehend beurteilt, in welchen Bereich es fällt. Dabei kann die Diskrepanz oder Ähnlichkeit eines Arguments überschätzt werden. Eine Einstellung, die als diskrepanter zur eigenen Sicht wahrgenommen wird, als sie es tatsächlich ist, unterliegt einem „Kontrast-Effekt“. Auf der anderen Seite können Meinungen, die der eigenen Haltung einer Person sehr ähneln, als ähnlicher wahrgenommen werden, als sie wirklich sind – ein „Assimilations-Effekt“ tritt auf. Botschaften, die zu einem Assimilations-Effekt führen, werden häufig von einer Meinungsänderung in die von ihnen vorgegebene Richtung gefolgt. Allerdings werden Anfangsmeinung und persuasive Botschaft in diesem Fall auch manchmal als zu ähnlich wahrgenommen, so dass die durchaus bestehende objektive Diskrepanz nicht mehr auffällt - was eine weitere Anpassung in Richtung der Nachricht scheinbar überflüssig macht. Nachrichten, die zu einem Kontrast-Effekt führen, können zudem in einigen Fällen einen „Bumerang-Effekt“ erzeugen. Dieser bedeutet eine Verschiebung der eigenen Haltung noch weiter weg von der persuasiven Botschaft wegen der großen Diskrepanz.

Wenn Bochner und Insko (1966) einen linear ansteigenden oder kurvilinearen Trend zwischen Diskrepanz und Einstellungsänderung feststellen, wie kann Yaniv (2004b) dann unter Bezugnahme auf eben diese Studie die Behauptung aufstellen, dass mit steigender Diskrepanz weniger Einstellungswandel einhergeht? Dies wird dadurch möglich, in dem er (Yaniv) den Umfang dieses Einstellungswandels, ähnlich wie auch bei der Bestimmung der Ratschlagsgewichtung vorgegangen wird, an der ihm vorausgegangenen Diskrepanz relativiert. Mit anderen Worten gelingt dies, wenn das Ausmaß der Einstellungsänderung und die Abweichungen zweier Meinungen zueinander ins Verhältnis gesetzt werden. Bei einer derartigen Transformation der von Bochner und Insko (1966) berichteten Ergebnisse ergibt sich ungeachtet der Quelle der persuasiven Botschaft ein monoton fallender Zusammenhang zwischen der Diskrepanz und der an ihr relativierten Einstellungsänderung. Andere Untersuchungen aus derselben Forschungs-Ära zeigen jedoch ein anderes Bild. Bei ihnen zeigt sich, dass das kurvilineare Verhältnis zwischen Diskrepanz und Einstellungsänderungen auch bei Transformation in solch eine relative Einstellungsänderung erhalten bleibt (Aronson et al.,

1963; Insko et al., 1966). Möglicherweise stellt der monoton fallende Zusammenhang zwischen den beiden Größen also nicht die Regel dar.

Obgleich das Maß der absoluten Einstellungsänderung nicht gleichzusetzen ist mit den an der Ratschlagsabweichung relativierten Maßen der Ratschlagsgewichtung, halte ich es dennoch für möglich, dass die Argumentation der SJT für Einstellungsänderungen auf Beratungssituationen übertragen werden kann. Demgemäß könnten Ratschläge, die weit von der ursprünglichen Schätzung liegen sollten, einen Kontrast-Effekt (Sherif, Taub & Hovland, 1958) hervorrufen, welcher eine geringe Gewichtung des Ratschlags zur Folge hat. Das heißt, wenn Ratschläge stark von der anfänglichen Schätzung eines Menschen abweichen, werden sie diskontiert, weil sie als übermäßig abweichend wahrgenommen werden und daher außerhalb eines bestimmten Akzeptanzbereichs von Werten liegen. Aus der Übertragung der Annahmen der SJT auf eine Beratungssituation würde weiterhin folgen – und dabei handelt es sich um den grundlegenden Unterschied zwischen meiner und Yanivs (2004b) Vorhersage – , dass Ratschläge, die einer anfänglichen Schätzung sehr stark ähneln, ebenfalls diskontiert werden; in diesem Fall jedoch auf Grund eines Assimilations-Effektes. Wenn ein Ratschlag also hinreichend nahe an einer Initialschätzung liegt, wird er unter Umständen vom Entscheidungsträger nicht als merklich unterschiedlich wahrgenommen – selbst wenn beide nicht numerisch identisch sind. Man stelle sich beispielsweise vor, eine Person schätzt eine bestimmte Entfernung auf 1000km. Im Anschluss erhält diese Person den Ratschlag einer anderen Person, welche dieselbe Entfernung auf 1020km schätzt. Diese beiden Schätzungen sind numerisch natürlich nicht identisch, dennoch ist dieser Unterschied für die meisten Menschen vernachlässigbar, da in diesem Fall durch den Assimilations-Effekt eine höhere psychische als objektive Ähnlichkeit zwischen der Initialschätzung und dem dargebotenen Ratschlag produziert werden könnte. Dadurch wird dieser nahe Ratschlag mit hoher Wahrscheinlichkeit als der Initialschätzung ähnlich oder mit ihr sogar nahezu identisch angesehen. Ist dies der Fall, so kann der Judge zu der Meinung gelangen, dass eine Anpassung der endgültigen Schätzung an den Ratschlag nicht notwendig ist.

Im Zusammenhang mit quantitativen Urteilen ist möglicherweise weniger von einem Akzeptanz- und Ablehnungsbereich im Sinne einer vorherrschenden Überzeugung auszugehen, als vielmehr von Plausibilitätsüberlegungen, die Menschen hinsichtlich der dargebotenen Ratschläge treffen. Demnach sollte bei Verwendung derlei Aufgaben möglicherweise eher von *„Latitude[s] of Plausibility“* (Wegener, Petty, Detweiler-Bedell, Blair & Jarvis, 2001, S. 64, Fußnote 1) oder *Implausibility* gesprochen werden. Wegener et al. (2001) halten diese (Un)Plausibilitätsbereiche sogar für geeigneter als die Berufung auf einen Ablehnungsbereich, um den Anfang der Abnahme von Einstellungsänderungen zu erklären.

Beim Vergleich der Argumentation der SJT und meiner daraus hervorgehenden Annahmen mit denen von Yaniv (2004b) ergibt sich, dass der Hauptunterschied im angenommenen Umgang (hinsichtlich der Gewichtung) mit nahen Ratschlägen liegt. Während die aus der SJT abgeleiteten Annahmen konstatieren, dass sowohl nahe als auch entfernte Ratschläge nur geringfügig gewichtet werden, nimmt Yaniv an, dass die Ratschlagsgewichtung mit zunehmender Distanz zwischen Initialschätzung und Ratschlag monoton abnimmt. Mit anderen Worten geht er davon aus, dass nahe Ratschläge stärker gewichtet werden als entfernte Ratschläge. Ich jedoch glaube, dass Yanivs Annahme dieses monotonen Trends mit Rücksicht auf die Gewichtung naher Ratschläge modifiziert werden muss. Der SJT folgend ist meine Annahme eine andere, nämlich dass nahe Ratschläge nicht stärker gewichtet werden als ferne Ratschläge. Ich vermute also, dass Ratschläge nicht mit steigender Entfernung zur Initialschätzung zunehmend geringer gewichtet werden, sondern eben auch nahe Ratschläge relativ geringfügig gewichtet werden, und zwar auf Grund ihrer sozial validierenden Wirkung, welche möglicherweise dadurch zu Stande kommt, dass ihre Ähnlichkeit zur Initialschätzung überschätzt wird. Zusammenfassend lässt sich demnach festhalten, dass sich zwei unterschiedliche Prozesse für die niedrige Gewichtung naher und ferner Ratschläge verantwortlich zeichnen: Während nahe Ratschläge ungewichtet bleiben, weil sie nicht als merklich unterschiedlich und dadurch vorwiegend als soziale Validierung der Ausgangsmeinung wahrgenommen werden, finden ferne Ratschläge nur wenig Beachtung, weil sie als wenig bis überhaupt nicht bestätigend empfunden werden, da sie außerhalb des Akzeptanzbereiches (bzw. Plausibilitätsbereiches) des Urteilenden liegen.

Die Befunde aus der Attitude Change-Forschung deuten also darauf hin, dass eine (umgekehrt) U-förmige Beziehung zwischen der Diskrepanz bzw. Distanz einer erhaltenen Information zu einem Ausgangspunkt und der Veränderung von diesem Ausgangspunkt zu einem Endpunkt herrscht. Das Ziehen solcher inhaltlicher Parallelen zwischen den Forschungsrichtungen der Einstellungsänderung und dem JAS erscheint dabei angemessen, da zwischen beiden Paradigmen große strukturelle Ähnlichkeiten vorliegen. So sind die Rollen der verschiedenen, am Kommunikationsprozess beteiligten Personen zumeist eindeutig verteilt bzw. zugewiesen. Der zu beeinflussenden Person (im JAS der Judge und in Attitude Change-Untersuchungen die oftmals als „Communicatee“ oder „Rezipient“ bezeichnete Person) stehen eine oder mehrere einflussnehmende Personen (im JAS der Advisor und in Attitude Change-Studien häufig der „Kommunikator“) gegenüber. In beiden Fällen ist die Änderung von einem Ausgangshin zu einem Endzustand in Abhängigkeit einer dazwischenliegenden Kommunikation, wie auch immer diese geartet sein mag, von Interesse. Dementsprechend bildet der Endzustand dieser Kommunikation in aller Regel ein Aggregat aus der eigenen Anfangs- und einer zusätzlich erhaltenen Einschätzung, Meinung, Information oder Einstellung ab.

Diese Ähnlichkeiten zwischen dem in der Attitude Change-Forschung verwendeten Paradigma und dem JAS lassen darauf schließen, dass sich diese umgekehrt U-förmige Beziehung möglicherweise auch im Zusammenhang mit Ratschlägen bzw. deren Gewichtung zeigen lässt. Yanivs Studie –und mit Abstrichen auch die von Minson et al.–, die die bisher einzige Grundlage zu Aussagen über den Zusammenhang zwischen Ratschlagsentfernung und –gewichtung darstellen, widersprechen dem natürlich teilweise. Dementsprechend gilt es, die von mir getätigten Annahmen im Rahmen der Experimente meiner Dissertation zu überprüfen.

3.2 Soziale Validierung als Aspekt der Ratschlagsnutzung

Obwohl Ratschlagsnutzung für gewöhnlich als Adjustierung der anfänglichen Schätzung in Richtung des erhaltenen Ratschlags definiert wird, gibt es – wie bereits im vorhergehenden Kapitel angedeutet – eine weitere Möglichkeit wie Ratschläge genutzt werden können, nämlich als eine Form der sozialen Validierung. So mutmaßen bereits

Gino und Moore (2007), dass ein Gewichtungskoeffizient (in ihrem Experiment das WOA) von 0 zwar keinerlei Adjustierung des Finalurteils in Richtung des erhaltenen Ratschlags anzeigt, dies jedoch im Falle einer Ähnlichkeit zwischen Initialschätzung und Ratschlag nicht zwangsläufig mit einer vollkommenen Nicht-Nutzung des Ratschlages gleichzusetzen ist. Der Judge könnte einen Ratschlag, der seinem Initialurteil ähnlich ist, stattdessen als Bestätigung für das Zutreffen des letzteren ansehen und dieses daraus folgend beibehalten haben. Alternativ sei es sogar möglich, dass Judges ihr Initialurteil nicht ändern, sondern lediglich ihr Vertrauen darin.

Das Vertrauen bzw. die subjektive Sicherheit von Judges hinsichtlich der von ihnen getroffenen Urteile und Entscheidungen im Rahmen des JAS wurde bereits verschiedentlich untersucht (siehe Bonaccio & Dalal, 2006). Zu diesem Zwecke erfolgt die Erhebung der subjektiven Sicherheit zumeist an zwei Zeitpunkten in den Untersuchungen, und zwar vor und nach Erhalt des Ratschlags, um Veränderungen im Laufe des Beratungsprozesses sichtbar machen zu können. Es konnte dadurch gezeigt werden, dass Menschen nach Erhalt eines Ratschlages grundsätzlich sicherer werden, also hinsichtlich ihres Finalurteils oder ihrer Finalentscheidung eine höhere subjektive –meist Likert-skalierte– Sicherheit angeben als hinsichtlich des Initialurteils oder der Initialentscheidung (Heath & Gonzales, 1995). Die Autoren mutmaßen, dass dieser Sicherheitszuwachs daraus resultiert, dass die Judges im Laufe des Beratungsprozesses und durch die Interaktion mit dem Ratgeber die Gelegenheit erhalten, eine Erklärung bzw. Argumentation und damit Untermauerung für ihr Urteil oder ihre Entscheidung zu entwickeln. Ein Forschungsschwerpunkt zur Urteilssicherheit der Judges liegt daher in der Untersuchung der Übereinstimmung von subjektiver Sicherheit und Akkuratheit der Urteile und Entscheidungen. In diesem Zusammenhang berichten Sniezek und Buckley (1995) von einer *Overconfidence* der Judges, sprich, dass diese gemessen an der Akkuratheit ihrer Finalentscheidungen eine zu hohe Sicherheit angeben. Weiterhin wurden verschiedene Faktoren untersucht, die das Sicherheits-Level nach dem Erhalt eines Ratschlags beeinflussen. Budescu und Rantilla (2000) stellten beispielsweise fest, dass die subjektive Urteilssicherheit der Judges höher ausfällt, wenn im Falle multipler Advisor eine hohe Übereinstimmung ihrer einzelnen Ratschläge besteht.

Soziale Validierung durch eine zweite oder mehrere zusätzliche Meinungen wurde überwiegend im Rahmen der Gruppenentscheidungsforschung erforscht. In diesem Rahmen konnte gezeigt werden, dass das Vertrauen in die Richtigkeit einer Gruppenentscheidung mit dem Ausmaß an sozialer Unterstützung für diese Entscheidung wächst. Mit anderen Worten haben meinungshomogene Gruppen und die sich darin befindlichen Individuen mehr Vertrauen in das Zutreffen ihrer Präferenz als Gruppen, in denen ein Dissens herrscht (Schulz-Hardt, Frey, Lüthgens & Moscovici, 2000; Schulz-Hardt, Jochims & Frey, 2002; Sniezek & Kolzow, 1992; zitiert nach Sniezek, 1992, S. 139-140). Wenn man zwei unabhängige Schätzungen, nämlich Initialschätzung und nachfolgenden Ratschlag, als eine Sonderform zweier unterschiedlicher Meinungen betrachtet, so könnte man erwarten, dass hier dieselbe Beziehung zwischen Unterschiedlichkeit und Urteilssicherheit hinsichtlich des Aggregats dieser beiden besteht.

Dies führt dann zu der Annahme, dass Ratschläge auch der sozialen Validierung dienen könnten, vor allem wenn sie der Initialschätzung eines Urteilers ähneln, denn in diesem Fall würden zwei unabhängige Quellen, nämlich Judge und Advisor, die gleiche „Meinung“ besitzen - dies ähnelt den geteilten Meinungen/Präferenzen in einer Gruppe (oder Dyade). Mit objektiv (und letztlich auch psychologisch) zunehmender Distanz zwischen Initialschätzung und Ratschlag wird eine soziale Validierung dann zunehmend unwahrscheinlich, wodurch sich die Urteilssicherheit allmählich verringern dürfte. Diese Annahme ist in Einklang mit Budescus (2006) Modell zur Urteilssicherheit bei Aggregation von Meinungen aus verschiedenen Quellen. Das Modell besagt, dass je mehr sich verschiedene Meinungen oder Urteile voneinander unterscheiden, desto weniger sicher sind Menschen hinsichtlich des aus ihnen hervorgehenden Meinungsaggregats.

Eine systematische Untersuchung der subjektiven Sicherheit in Abhängigkeit der Entfernung zwischen dem Initialurteil des Judges und einem im Anschluss erhaltenen einzelnen Ratschlag gibt es meines Wissens bisher nicht. Auch Yaniv (2004b) und Minson et al. (2011) haben in ihren Untersuchungen keine Urteilssicherheit der Probanden erhoben. Dementsprechend gibt es auch noch keine empirische Untersuchung, die sich in diesem Zusammenhang dem Ratschlag als Quelle sozialer Validierung quan-

titativ widmet. Cross (2000; zitiert nach Cross, Borgatti und Parker, 2001) stellte im Rahmen einer qualitativen Untersuchung fest, dass die soziale Validierung eine Funktion des Einholens von Ratschlägen erfüllt. In einer Befragung von Managern kristallisierte sich diese Funktion bzw. Qualität von Ratschlägen (neben einer Reihe von anderen) heraus, welche sich konkret laut den Berichten der befragten Manager zum Beispiel durch den Zuspruch oder die Bestätigung eines bereits eingeschlagenen Weges innerhalb eines zu bearbeitenden Projektes äußerte. Weiterhin berichteten sie (die Manager), dass diese validierenden Interaktionen ihnen geholfen haben, ein gewisses Vertrauen bzw. eine gewisse Sicherheit hinsichtlich ihrer Konzepte oder Pläne zu entwickeln, die sie dadurch in Folge und anderen Kontexten noch überzeugter vertreten konnten (Cross & Sproull, 2004). Diese Forschungsergebnisse liefern damit einen Hinweis darauf, dass die soziale Validierung einen Nutzen für die validierten Personen aufweist. Daraus könnte geschlossen werden, dass auch sie einen Aspekt der Ratschlagsnutzung verkörpert. Falls dies zutrifft, dürfte der Begriff der Ratschlags*nutzung* nicht vorschnell vollständig mit dem der Ratschlags*gewichtung* gleichgesetzt und in Folge erstere nicht vollständig durch letztere operationalisiert werden. Dies wird auch deutlich, wenn meine Ausführungen aus Kapitel 3.1 hier herangezogen werden. Würden hier Ratschlagsnutzung und –gewichtung als äquivalent angesehen werden, so gelangte man zu dem Fazit, dass nahe und auch ferne Ratschläge nur sehr wenig bis überhaupt nicht genutzt werden. Folgt man jedoch auch den Ausführungen zur sozialen Validierung, so wird deutlich, dass die Wirkung von nahen und entfernten Ratschlägen auf die Probanden sehr unterschiedlich sein dürfte. So werden nahe Ratschläge zwar nur geringfügig gewichtet, mit hoher Wahrscheinlichkeit allerdings als Quelle der Absicherung genutzt, während bei entfernten Ratschlägen hiervon nicht auszugehen ist, das heißt hier sollte eine geringe Gewichtung tatsächlich ein Zeichen geringer Nutzung sein.

Zusammenfassend lässt sich also festhalten, dass soziale Validierung möglicherweise einen weiteren, die Ratschlagsgewichtung ergänzenden Aspekt der Ratschlagsnutzung darstellt. Als Indikator für das Ausmaß der empfundenen Validierung kann dabei die subjektive Urteilssicherheit bzw. deren Veränderung von Initial- zu Finalurteil herangezogen werden. Auch diese Sicherheit ist höchstwahrscheinlich abhän-

gig von der Unterschiedlichkeit zwischen Initialurteil und Ratschlag; und zwar dahingehend, dass sie bei wachsender Distanz zwischen diesen beiden Werten immer kleiner ausfällt. Dementsprechend werden also vermutlich ausschließlich nahe bzw. der eigenen Meinung sehr ähnliche Ratschläge als soziale Validierung empfunden. In sehr wenigen Untersuchungen (mit Ausnahme von z.B. Sniezek & Buckley, 1995) wurden bisher sowohl ein Maß für die Ratschlagsgewichtung als auch gleichzeitig die Urteilssicherheit erhoben. Selbst wenn dies jedoch der Fall war, wurden das Gewichtungsmaß und die Sicherheit nicht unbedingt im Hinblick auf ihre sich ergänzende Funktion für die Ratschlagsnutzung analysiert. Es erscheint mir demnach sinnvoll, aber auch notwendig, die subjektive Sicherheit von Personen in Abhängigkeit der Unterschiedlichkeit oder auch Gleichheit ihrer Eingangsschätzungen und den Empfehlungen eines Ratgebers systematisch zu überprüfen, um somit vielleicht das vorherrschende Verständnis des Begriffs der Ratschlagsnutzung erweitern zu können.

Die Überprüfung meiner Forschungsfragen erfolgt in einer Serie von vier Experimenten, die dem Muster des Judge-Advisor-Paradigmas folgen. In allen Experimenten bearbeiten die Versuchseilnehmer eine Reihe von quantitativen Urteilsaufgaben, nämlich Entfernungs- oder Kalorienschätzungen. Experiment 1 untersucht die Beziehung zwischen Ratschlagsentfernung als unabhängiger Variable und Ratschlagsgewichtung sowie Veränderung der Urteilssicherheit als abhängigen Variablen. In Experiment 2 werden die Ergebnisse aus Experiment 1 unter Verwendung eines weiteren Aufgabentyps repliziert. Experiment 3 untersucht die Ratgeberkompetenz als potenziellen Moderator des Distanzeffektes. Abschließend wird in Experiment 4 überprüft, ob die in den Experimenten 1 bis 3 gefundenen Effekte von der Art der Operationalisierung der Ratschlagsentfernung abhängen.

4. Experiment 1

4.1 Zielsetzung und Hypothesen

Experiment 1 beschäftigt sich mit der Art des Verhältnisses zwischen der Entfernung von Initialschätzung und Ratschlägen einerseits und der Ratschlagsgewichtung sowie der Veränderung der subjektiven Urteilssicherheit andererseits. Es soll den ersten Nachweis dafür liefern, dass sich die von mir aufgestellten Überlegungen hinsichtlich des Zusammenhangs dieser Variablen empirisch untermauern lassen. Gemäß meinen voran gegangenen theoretischen Ausführungen und den daraus abgeleiteten Forschungsfragen postuliere ich folgende Hypothesen:

Hypothese 1: Die Ratschlagsgewichtung ist eine quadratische (umgekehrt U-förmige) Funktion der Distanz zwischen Initialschätzung und Ratschlag, das heißt Ratschläge werden weniger stark gewichtet werden, wenn sie entweder sehr nah an oder sehr weit von der Initialschätzung des Judges entfernt sind, im Vergleich zu Ratschlägen, deren Entfernung zur Initialschätzung moderat ausfällt.

Meinen Ausführungen zur sozialen Validierung durch nahe Ratschläge folgend, lässt sich postulieren, dass die Urteilssicherheit der Judges bezüglich der aus Initialschätzung und Ratschlag hervorgehenden kombinierten Schätzungen eine monoton fallende Funktion der Ratschlagsdistanz sein sollte. In anderen Worten, während Ratschlagsnutzung im Sinne ihrer Gewichtung ähnlich für sehr nahe und sehr ferne Ratschläge ausfallen sollte, gehe ich davon aus, dass nahe und ferne Ratschläge sich hinsichtlich der mit ihnen verbundenen Urteilssicherheit grundsätzlich unterscheiden. Dies ist in Hypothese 2 wie folgt formuliert:

Hypothese 2: Je näher Ratschläge an der Initialschätzung liegen, desto stärker wird die Urteilssicherheit zwischen Initial- und Finalurteil ansteigen.

4.2 Methode

4.2.1. Stichprobe und Design des Experimentes

27 Studenten (*M* = 21.19 Jahre, *SD* = 2.35) nahmen an Experiment 1 teil. Die Stichprobe bestand aus 19 (70%) weiblichen und acht männlichen Teilnehmern. Für die Teilnahme erhielten sie entweder eine Versuchspersonenstunde oder eine monetäre Vergütung von 6€. Um sicherzustellen, dass die Teilnehmer eine grundlegende Motivation zu akkuraten Schätzungen aufweisen, wurden sie informiert, dass die besten fünf von ihnen (in Bezug auf die Genauigkeit ihrer Schätzungen) Kino-Gutscheine im Wert von 15 Euro erhalten würden.

Die Experimente 1 bis 4 wurden in den Laborräumen des Instituts für Psychologie durchgeführt. Den Teilnehmern wurde zuvor bei der Akquise mitgeteilt, dass sie in dem Experiment Schätzaufgaben zu bearbeiten hätten. Obwohl mehrere Personen zu einem Termin eingeladen wurden, verlief die Durchführung non-interaktiv. Die Teilnehmer saßen getrennt voneinander in den Laborräumen und erhielten alle erforderlichen Instruktionen via PC.

Experiment 1 weist ein einfaktorielles Design mit dem 15-stufigen Innersubjektfaktor *Ratschlagsdistanz* (5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%) auf. Die Durchführung des Experiments erfolgte computerbasiert unter Verwendung der Experimentalsoftware Presentation® (Version 13.0 – 14.8, www.neurobs.com), mit der alle vier Experimente programmiert wurden. Die entsprechenden Programmcodes für die Experimente 1 bis 4 befinden sich im digitalen Anhang dieser Arbeit.

4.2.2 Ablauf

Zu Beginn des Experimentes wurde allen Versuchseilnehmern die Rolle des Entscheidungsträgers, also des Judges, zugeteilt. Das heißt, alle Probanden wurden darüber informiert, dass sie für das Treffen einer initialen sowie finalen Schätzung zuständig seien. Ihnen wurde weiterhin mitgeteilt, dass sie vor dem Treffen eines Finalurteils einen Ratschlag erhalten würden. Zu diesem Zweck würde zu Beginn des Experiments

zufällig aus einem Pool von 100 Personen, die angeblich zu einem früheren Zeitpunkt an einer Untersuchung teilgenommen hatten, bei der dieselben Schätzaufgaben verwendet worden waren, eine Person gezogen, deren Schätzungen den jetzigen Versuchsteilnehmern als Ratschläge zur Verfügung stünden. Diese angeblichen früheren Versuchsteilnehmer seien zudem an Hand der Akkuratheit ihrer Schätzungen in eine Rangfolge gebracht worden, bei welcher der erste Rang die am genauesten schätzende Person und Rang 100 die am wenigsten akkurat schätzende Person kennzeichnete. Zur Erhöhung der Augenscheinvalidität der Zufallsauswahl wurde den Probanden ein vermeintlicher Zufallsgenerator präsentiert. Dazu erfolgte ein schneller und wiederholter Durchlauf der Zahlen von 1 bis 100, welcher angeblich durch Drücken der Leertaste auf der Tastatur beendet werden konnte. Den Teilnehmern wurde dann gesagt, sie würden Ratschläge von der Person, deren Rang mit der Zahl beim Stoppen des Generators übereinstimmt, erhalten. In Wahrheit kam dieser Zufallsgenerator nach Betätigung der Leertaste durch die Probanden jedoch immer bei „7“ zum Stillstand, woraufhin ihnen ein vermeintlicher Ratgeber zugordnet wurde, der demgemäß den siebten Rang von 100 angeblichen früheren Probanden erreicht hatte und dessen Kompetenz daher als recht hoch anzusehen ist. Die Wahl eines derart kompetenten Ratgebers erfolgte auf Grund der Erkenntnis aus vorhergehenden Untersuchungen, dass recht kompetente Ratgeber in etwa zu 50% gewichtet werden (z.B. Harvey & Fischer, 1997; Soll & Mannes, 2011; Yaniv & Kleinberger, 2000). Dieses Ausmaß der Gewichtung erlaubt einen genügend großen Spielraum für abnehmende und steigende Gewichtungen von Ratschlägen in Abhängigkeit ihrer Distanz. Bei gleich- bis weniger kompetenten Ratgebern im Vergleich zum Judge sollte die Gewichtung hingegen gemäß der bisherigen Literatur bei etwa 30% liegen (vgl. Harver & Fischer, 1997), was weniger Spielraum für Anpassungen der Gewichtung nach unten ließe.

Wie an Hand der Manipulation der vermeintlichen Ratgeberkompetenz bereits zu erahnen ist, handelte es sich bei den dargebotenen Hilfestellungen nicht um echte Ratschläge. Stattdessen wurden die Ratschläge jedes Mal durch die Addition oder Subtraktion bestimmter aus den randomisierten Abweichungen resultierender Werte zur Initialschätzung gebildet. Mit solch mechanisch konstruierten Ratschlägen lässt sich sehr präzise überprüfen, welchen systematischen Einfluss kontinuierlich an Abstand

zunehmende Ratschläge auf eine Finalschätzung ausüben (siehe auch Yaniv, 2004b). Der Abstand zwischen der Initialschätzung des Judges und den erhaltenen Ratschlägen wurde als within-subject-Variable manipuliert, indem die Initialschätzung durch Addition oder Subtraktion eines bestimmten Prozentsatzes von ± 5% bis ± 75% in 5%-Schritten lineartransformiert wurde. Die Reihenfolge, in der die Abweichungen auftraten, war vor Beginn des Experiments randomisiert worden, wurde jedoch für alle Probanden gleich gehalten, um Fehlervarianz zu reduzieren. Das Programm enthielt zudem einen Sicherungsmechanismus, um zu verhindern, dass Probanden die Zusammenhänge zwischen eigener Initialschätzung und dem dargebotenen Ratschlag erkennen können. Zu diesem Zweck wurde eine Zufallszahl zwischen 500km und 4000km als „scheinbar authentische“ Schätzung generiert, wenn der errechnete Ratschlag kleiner als 25km oder aber größer als 7500km ausgefallen wäre (was in der Regel das Ergebnis eines Tippfehlers bei der Initialschätzung war).

Experiment 1 bestand aus 60 Durchgängen, in denen die Teilnehmer die Aufgabe erhielten, die Entfernung zwischen jeweils zwei Europäischen Hauptstädten zu schätzen. Jeder Versuch wies entsprechend dem Judge-Advisor-Paradigma drei Phasen auf. In der ersten Phase wurden den Befragten die Namen der zwei Städte eingeblendet, deren Entfernung es anschließend zu schätzen galt. Nachdem eine initiale Entfernungsschätzung erfolgt war, wurden die Probanden dazu aufgefordert anzugeben, wie sicher sie sich hinsichtlich dieser ersten Schätzung sind (Initialsicherheit). Dies geschah auf einer 7-Punkt-Likert-Skala mit den Endpunkten 1 (= sehr unsicher) und 7 (= sehr sicher). In der zweiten Phase eines Versuchsdurchgangs wurde den Probanden der hinsichtlich seiner Abweichung von der Initialschätzung manipulierte Ratschlag dargeboten. Im Anschluss sollten die Versuchsteilnehmer in der dritten Durchgangsphase eine Finalschätzung zur Entfernung der zuvor genannten Hauptstädte abgeben und danach erneut ihre subjektive Sicherheit hinsichtlich dieser Schätzung angeben (Finalsicherheit).

Zur Vermeidung möglicher stereotyp-bezogener Auswirkungen, welche auf Geschlechtsunterschiede zwischen Ratgeber und Judge zurückzuführen sind, wurde das Geschlecht von Judge und Advisor aufeinander abgestimmt. Weibliche Versuchsteil-

nehmer erhielten Ratschläge von „Christine“, männliche von „Christian“. Die Teilnehmer wurden darüber informiert, dass dies nicht die wirklichen Namen der Ratgeber waren, sondern dass diese Namen willkürlich gewählt wurden, um ihre Anonymität zu gewährleisten.

Zur besseren Verständlichkeit ist der Versuchsablauf in Abbildung 1 noch einmal überblicksartig aufgeführt.

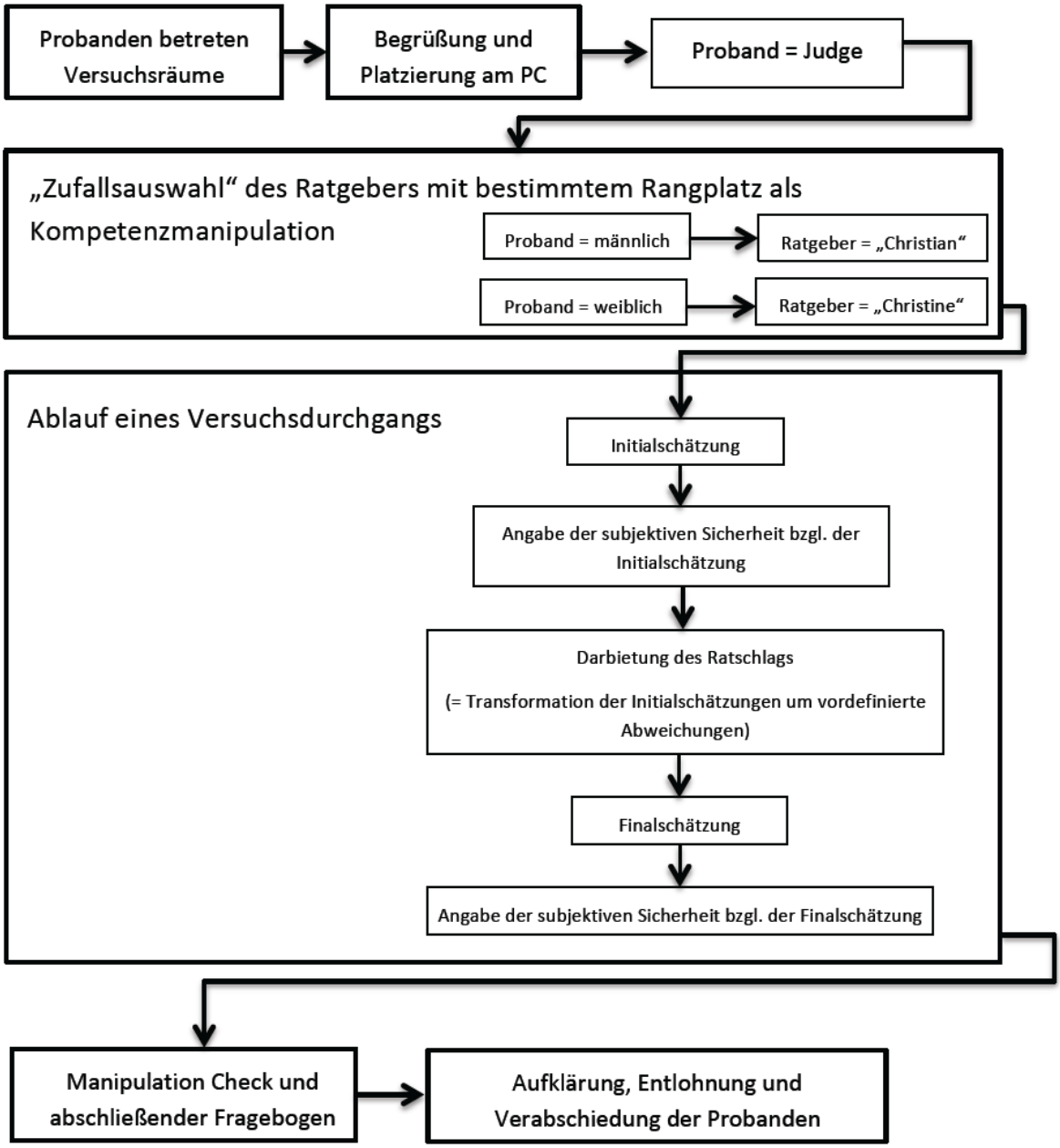

Abbildung 1

Schematischer Ablauf der Experimente 1 bis 4

4.3 Ergebnisse und Diskussion

4.3.1 Analyse möglicher Störvariablen

Als potentielle Störvariablen wurden Einflüsse des Alters der Probanden auf die mittlere Ratschlagsgewichtung und eventuelle Effekte des Faktors Geschlecht bei Einbezug in die Hauptanalyse wie in Kapitel 4.3.2 und 4.3.3 beschrieben überprüft. Zu diesem Zwecke wurden für alle Probanden zunächst die beiden Kennzahlen für die abhängigen Variablen Ratschlagsgewichtung (AT; Kapitel 4.3.2) und Urteilssicherheit (Finalsicherheit – Initialsicherheit; Kapitel 4.3.3) pro manipulierter Ratschlagsabweichung berechnet. Es wurden nur AT-Werte im Bereich von -1.5 bis +1.5 in die statistische Analyse miteinbezogen. Zwar sind AT-Werte, die oberhalb oder unterhalb dieses Wertebereichs liegen, durchaus denkbar und möglich; in der Mehrzahl der Untersuchungen, die das AT- oder vergleichbare Maße verwenden, liegen die empirisch ermittelten jedoch zwischen 0 und 1 (siehe Kapitel 2.2). Um aber dennoch nicht von vornherein jegliche Anpassung der Finalschätzung über den Range aus Initialschätzung und Ratschlag hinaus oder aber auch ihm entgegen vollständig auszuschließen –wofür es überdies auch gar keine theoretische Begründung gäbe–, wurde der gültige Wertebereich etwas liberaler gewählt. Zusätzlich wurden aber auch solche AT-Werte, und die im selben Trial angegebenen Urteilssicherheiten bzw. die Veränderung derselben, von der Analyse ausgeschlossen, wenn der dazugehörige computergenerierte Ratschlag nicht die ausgewiesene Abweichung von der Initialschätzung aufwies. Dies geschah, wenn der Computer statt einer Lineartransformation als Ratschlag eine Zufallszahl generierte, weil der unter Anwendung der vorgegebenen Abweichungen errechnete Ratschlag zu hoch oder zu niedrig ausgefallen wäre, um augenscheinlich plausibel zu sein. Wenn die Initialschätzung eines Versuchsteilnehmers also beispielsweise 5000km betragen hätte und durch Zufall in diesem Versuchsdurchgang die +75%-Abweichung angewendet hätte werden müssen, hätte sich der Ratschlag bei Anwendung der Lineartransformation auf unwahrscheinlich hohe 8750km belaufen. Stattdessen wurde daher in so einem Fall eine Zufallszahl zwischen 500 und 4000 dargeboten. Insgesamt mussten so in diesem Experiment 90 von 1620 Versuchsdurchgängen ausgeschlossen werden. Mit einer Rate von knapp 6% Ausschlüssen kommt damit ein vergleichsweise üblicher Wert (vgl. Gino, 2008; Gino et al., 2009) zustande. Im Anschluss wurde mit

den bleibenden AT-Werten die mittlere Ratschlagsgewichtung eines jeden Probanden berechnet und mit dem Alter der Probanden korreliert. Es zeigte sich kein signifikanter Zusammenhang zwischen diesen beiden Variablen, $r(27) = -.21$, $p = .29$. Ebenso erwies sich auch der Zusammenhang zwischen Alter der Probanden und der mittleren Veränderung der Urteilssicherheit als nicht systematisch, $r(27) = .11$, $p = .59$. Eine Aufnahme des Geschlechts der Probanden als Zwischensubjektfaktor zusätzlich zum Innensubjektfaktor Ratschlagsabweichung in die Haupt-ANOVA zeigte weder einen signifikanten Haupteffekt auf die AV Ratschlagsgewichtung, $F(1, 25) = 3.03$, $p = .09$, $\eta^2 = .11$, noch auf die AV Urteilssicherheit, $F(1, 25) = 2.88$, $p = .102$, $\eta^2 = .103$. Bei weiblichen Probanden zeigten sich allerdings tendenziell sowohl eine durchschnittlich höhere Ratschlagsgewichtung, als auch eine höhere Veränderung der Urteilssicherheit im Laufe der Versuchsdurchgänge. Bei keiner der beiden AVn ergab sich eine signifikante Interaktion zwischen Inner- und Zwischensubjektfaktor, beide Fs < 1.8, beide ps > .1, beide $\eta^2 <$ .07. Abschließend wurde noch überprüft, ob die Richtung der Ratschlagsabweichung einen systematischen Effekt auf die Ratschlagsgewichtung ausübt. Dies ist notwendig, weil bei den verwendeten Schätzaufgaben naturgemäß eine untere Grenze von 0, jedoch keine obere Grenze vorhanden ist. Eine ANOVA mit der *Richtung der Abweichung* als Innersubjektfaktor ergab keinen signifikanten Haupteffekt der Richtung der Abweichung, $F(1, 26) = 0.33$, $p = .57$, $\eta^2 = .01$. Die Höhe der Ratschlagsgewichtung ist demnach unabhängig davon, ob ein dargebotener Ratschlag höher oder niedriger als die Initialschätzung eines Probanden ausfällt.

4.3.2 Ratschlagsgewichtung

Pro Versuchsteilnehmer wurde für jeden einzelnen Versuchsdurchgang ein AT-Wert berechnet. Wie bereits erwähnt, wurde jede positive und negative Ratschlagsabweichung im Versuch zweimal angewandt. Die vier AT-Werte, die so jeweils pro Versuchsteilnehmer und (betragsmäßiger) Ratschlagsabweichung entstanden sind, wurden im Anschluss gemittelt. Wie eingangs in Kapitel 4.3.1 beschrieben, wurden hierfür jedoch nur AT-Werte zwischen -1.5 und + 1.5 und solche, die in Trials mit einer echten Distanzmanipulation entstanden sind eingeschlossen. Eine Unterscheidung der AT-Werte nach positiven und negativen Abweichungen erfolgte nicht, da kein Nachweis dafür existiert, dass die Richtung der Abweichung einen systematischen Effekt auf die

Ratschlagsgewichtung ausübt (vgl. Kapitel 4.3.1 und Schultze, 2010). Es ergab sich auf diesem Wege pro Proband und Ratschlagsabweichung ein (gemittelter individueller) AT-Wert, welcher als AV in die Datenmaske zur Berechnung der nachfolgenden Varianzanalyse mit Messwiederholung eingetragen wurde.

Im Anschluss wurden diese AT-Werte mittels einer Varianzanalyse mit Messwiederholung (hiernach rm-ANOVA) analysiert. Dabei wurden im Speziellen die Kontraste des Innersubjektfaktors Ratschlagsdistanz überprüft bzw. eine Trendanalyse vorgenommen. Wenn, wie bei Yaniv (2004b), Ratschläge umso weniger gewichtet werden, je weiter sie von der Initialschätzung abweichen, dann sollte der lineare Kontrast Signifikanz erzielen. Wenn, wie von mir angenommen, dagegen Ratschläge auch relativ gering gewichtet würden, wenn sie sehr nahe an der Initialschätzung liegen, sollte (auch) der quadratische Kontrast signifikant werden. Die Ergebnisse der rm-ANOVA zeigen, dass die Ratschlagsdistanz insgesamt einen systematischen Einfluss auf die Ratschlagsgewichtung hat, $F(7.34, 190.92)$[2] $= 6.498$, $p < .001$, $\eta^2 = .20$. Der grafische Zusammenhang ist zudem durch einen quadratischen Kontrast gekennzeichnet, $F(1, 26) = 23.38$, $p < .001$, $\eta^2 =. 47$. Hypothese 1 kann damit angenommen werden. Der lineare Kontrast erlangt ebenfalls Signifikanz, $F\ (1, 26) = 15.06$, $p = .001$, $\eta^2 = .37$, was darin bergründet sein dürfte, dass die Gewichtung der Ratschläge mit hohen Abweichungen nicht auf das niedrige Niveau sehr naher Ratschläge zurückfällt.

[2] Ein Mauchley-Test zeigte, dass die Annahme der Sphärizität in der überwiegenden Mehrheit der Analysen verletzt war. Daher werden in den meisten ANOVAs mit Messwiederholung in allen vier Experimenten korrigierte Freiheitsgrade (nach Grennhouse-Geisser) berichtet.

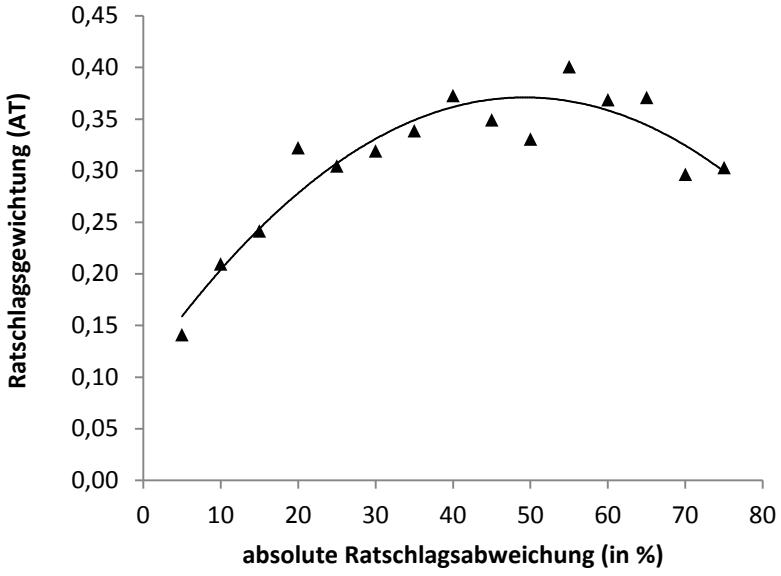

Abbildung 2

grafische Beziehung zwischen Ratschlagsdistanz (in %) und mittlerer Ratschlagsgewichtung in Experiment 1

An dieser Stelle ist zu beachten, dass in der Bezugsstudie von Yaniv (2004b) nicht das AT-Maß als Indikator für die Ratschlagsgewichtung berechnet wurde, sondern das von Yaniv selbst entwickelte Weight of Advice (WOA). Wie in Kapitel 2.2 ausgeführt, besteht der wesentliche Unterschied zwischen den beiden Maßen darin, dass das AT negative Werte annehmen kann und das WOA nicht. Demgemäß kann es unter den Umständen, dass Menschen ihre Finalschätzung entgegen des Ratschlags anpassen, bei Verwendung des WOAs zu einer Überschätzung der Ratschlagsgewichtung kommen. Um dies zu vermeiden, habe ich in meiner Dissertation auf das AT zurückgegriffen. Es ist jedoch denkbar, dass die unterschiedliche Berechnung der beiden Maße zu systematischen Unterschieden hinsichtlich des Distanzeffektes geführt hat, sodass die Unterschiede zwischen Yanivs (2004b) und meinen Befunden eine Folge der Verwendung unterschiedlicher Gewichtungsmaße darstellt. Für Experiment 1 wurde daher die Analyse zum Distanzeffekt bei der Ratschlagsgewichtung noch einmal unter Verwendung des WOA berechnet. Der Gesamtmittelwert des WOA (M = .32, SD = .20), also die durchschnittliche Ratschlagsgewichtung unterschied sich dabei nicht signifikant von der, die mittels des AT berechnet wurde (M = .31, SD = .21), $t(26) = -.275$, p = .79, d = .16. Weiterhin ergab sich auch für das WOA als abhängige Variable ein signifikanter Distanzeffekt, $F(7.85, 204.18) = 6.595$, $p < .001$, $\eta^2 = .20$. Der Zusammenhang

zwischen Ratschlagsentfernung und Ratschlagsgewichtung, abgebildet mit dem WOA, wurde sowohl durch einen linearen, $F(1, 26) = 16.65$ $p < .001$, $\eta^2 = .39$, als auch durch einen quadratischen Kontrast, $F(1, 26) = 21.35$, $p < .001$, $\eta^2 = .45$, beschrieben. Daraus geht deutlich hervor, dass die gefundenen Zusammenhänge nicht ein Produkt der verwendeten Gewichtungskennzahlen darstellen. Ebenso spricht die Kongruenz der beiden Werte für den eingangs berichteten Befund, dass die Mehrzahl der Finalurteile im Bereich zwischen Initialschätzung und Ratschlag liegen (Gino et al., 2009; Soll & Larrick, 2009; und siehe Kapitel 2.2), weswegen sie sich im Mittel trotz unterschiedlicher Berechnung nicht wesentlich voneinander unterscheiden.

Insgesamt gewichteten die Versuchsteilnehmer aus Experiment 1 erhaltene Ratschläge im Mittel also zu .31 (AT-Wert) – ein vergleichsweise typischer Wert der Adjustierung (vgl. Harvey & Fischer, 1997). Die mittleren Gewichtungen für die einzelnen Ratschlagsentfernungen variieren dabei stark, was zur Entstehung der postulierten umgekehrten U-Form des grafischen Zusammenhangs zwischen Ratschlagsdistanz und –gewichtung führt. Seit Soll und Larrick (2009) ist jedoch bekannt, dass Probanden häufig nicht das Gewichtungsmaß an den Ratschlag anlegen, welches durch berechnete Mittelwerte angezeigt wird, das heißt, der errechnete AT von .31 kommt höchstwahrscheinlich nicht dadurch zu Stande, dass sich die Versuchsteilnehmer aus Experiment 1 pro Versuchsdurchgang überwiegend zu etwa 30% dem Ratschlag angenähert haben. Stattdessen haben sie vermutlich vorwiegend zwischen zwei Alternativen zur Generierung eines finalen Urteils gewählt: Entweder sie entschieden sich für eine der dargebotenen Meinungen, Informationen oder eines der dargebotenen Urteile, inklusive des eigenen Standpunktes („*Choosing*“, nach Soll & Larrick, 2009), oder sie mittelten die vorhandenen Angebote, um zu einem Endurteil zu gelangen („*Averaging*“, nach Soll & Larrick, 2009). Durch die unterschiedlich häufige Anwendung der einen oder der anderen Strategie, das heißt verschieden häufig vorkommenden AT-Werten von 0, 0.5 oder auch 1, entstehen dann verschiedene mittlere Gewichtungswerte. Soll und Larrick (2009) fanden in ihrem Experiment, dass die relativen Häufigkeiten der Nullgewichtungen, 50%- und teilweise auch 100%-Gewichtungen am höchsten ausfielen, wodurch eine drei-gipflige Verteilung der Gewichtung entstand.

Um Aufschluss über die verwendeten Gewichtungsstrategien der Probanden in meinen Experimenten zu erhalten, wurde die jeweilige Gewichtungsverteilung analysiert. Zu diesem Zweck wurden die errechneten AT-Werte pro Person und Versuchsdurchgang entsprechend ihres Wertes sieben möglichen Gewichtungskategorien zugeordnet wie sie auch bei Soll und Larrick (2009) verwendet wurden (Nullgewichtung, > 0- bis .20-Gewichtung, ≥ .20- bis < .40-Gewichtung, ≥ .40- bis < .61-Gewichtung, ≥ .61- bis < .80-Gewichtung, ≥ .80- bis < 1-Gewichtung und 1- bzw. Vollgewichtung des Ratschlages) und anschließend deren Auftretenshäufigkeit bestimmt. In Abbildung 3 ist die Verteilung der Häufigkeiten der einzelnen Gewichtungskategorien (ungeachtet der Ratschlagsdistanz) abgebildet. Die Graphik ähnelt dabei stark den von Soll und Larrick (2009) dargestellten, mit Ausnahme des fehlenden dritten Gipfels bei der Vollgewichtung. Probanden in meinem ersten Experiment gewichteten Ratschläge eher selten zu 100%. Stattdessen behielten sie überwiegend ihre eigene Initialschätzung bei oder bildeten den Mittelwert (ausgedrückt durch die Kategorie der ≥ .40- bis < .61-Gewichtung) aus Eigenschätzung und Ratschlag, wodurch eine zwei-gipflige Verteilung der Gewichtung resultiert. Relativ häufig kamen auch Gewichtungen im Bereich von .20 bis knapp .40 vor, welche die von Harvey und Fischer (1997) konstatierte Grundgewichtung widerspiegeln dürften. Wenn also nicht die eigene Schätzung beibehalten oder mit Hilfe des Ratschlags ein Mittelwert gebildet wurde, wurde der Ratschlag zumindest in einem „Höflichkeitsausmaß“ gewichtet. Erwähnenswert ist hierbei jedoch, dass diese Gewichtungskategorie eben nicht die am häufigsten vorkommende war, was auf Grund des Gesamt-AT-Mittelwertes jedoch suggeriert hätte werden können.

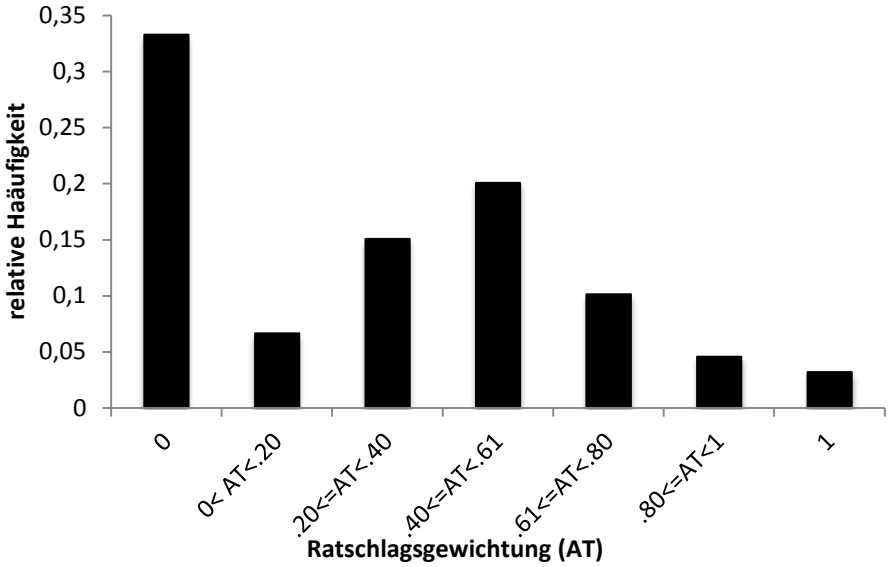

Abbildung 3

Verteilung der Ratschlagsgewichtungen in Experiment 1

Zum genauen Verständnis des Zustandekommens der mittleren Gewichtungen pro Ratschlagsentfernung und damit der Form des grafischen Zusammenhangs zwischen Entfernung und Gewichtung wurde für Experiment 1 auch die Verteilung der Gewichtungen in Abhängigkeit der ihnen zu Grunde liegenden Ratschlagsdistanzen analysiert. Dafür wurde die Auftretenshäufigkeit der einzelnen Kategorien pro manipulierter Ratschlagsentfernung ausgezählt und für betragsmäßig gleiche Entfernungen zusammengefasst. In Abbildung 4 sind die Verteilungen der relativen Häufigkeiten der Gewichtungen pro absoluter Ratschlagsabweichung dargestellt.

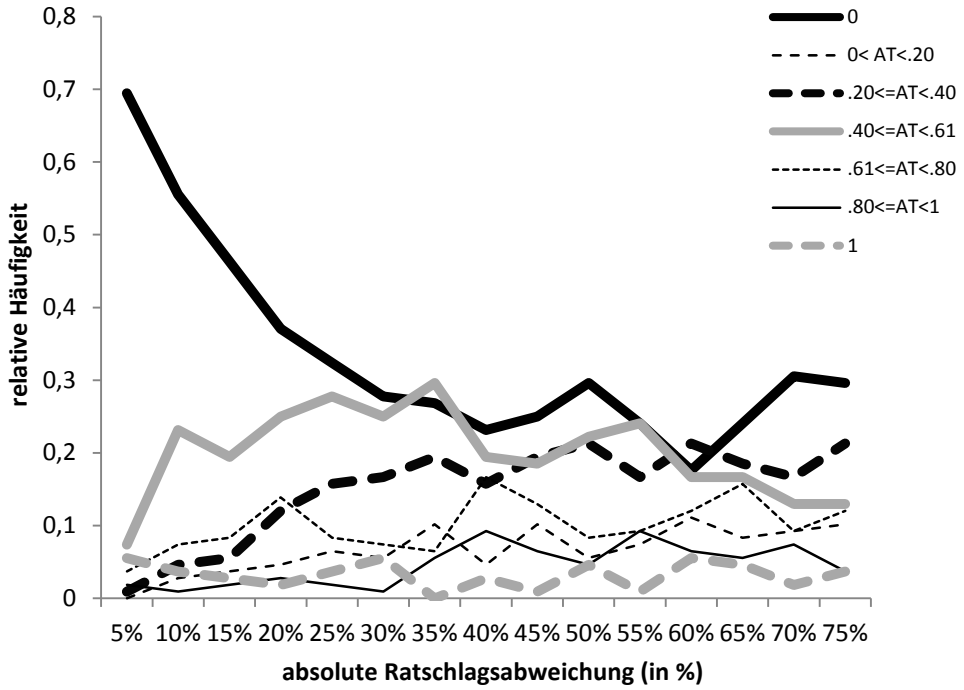

Abbildung 4

relative Häufigkeiten der Ratschlagsgewichtungen in Abhängigkeit der absoluten Ratschlagsentfernung in Experiment 1

Wie aus dieser Grafik ebenfalls ersichtlich ist, gewichteten die Probanden in Experiment 1, wie bereits erwähnt, überwiegend ihre eigenen Schätzungen und diskontierten erhaltene Ratschläge in großem Umfang. Die Kurve der Nullgewichtung liegt mit wenigen Ausnahmen fast durchgängig über den anderen Gewichtungskurven und kennzeichnet damit den relativ hohen Anteil an Nullgewichtungen gegenüber anderen Gewichtungsbeträgen pro Ratschlagsdistanz. Wenn Probanden Ratschläge nicht vollständig diskontierten, schien ihre bevorzugte Gewichtungsstrategie überwiegend darin zu liegen, durch Mittelung der eigenen Schätzung und dem erhaltenen Ratschlag zu einem Finalurteil zu gelangen. Fast über alle Distanzen hinweg ist die Gewichtungskurve der ≥40- bis <.61-Kategorie die zweithäufigste. Aus der Grafik geht weiterhin hervor, dass sich die Auftretenshäufigkeiten der einzelnen Gewichtungen in Abhängigkeit der Ratschlagsabweichung verändern. Die Gewichtungsverteilungen zeigen dabei, was auch bereits schon in den Daten von Soll und Larrick (2009) zum Vorschein kam: Der Mittelwert der Gewichtungen pro absoluter Ratschlagsdistanz (siehe Abbildung 2) stellt keinen typischen Wert dar, sondern entsteht mit größerer Wahrscheinlichkeit aus einer Kombination verschieden häufig auftretender Gewichtungsstrategien. Bei

diesen Strategien handelt es sich dann, wie bereits erwähnt, überwiegend um die Mittelung (*Averaging*) von Initialschätzung und Ratschlag (über 40%- bis einschließlich 60%-Gewichtung) und die Wahl zwischen der Beibehaltung der eigenen Initialschätzung (Nullgewichtung) oder der vollständigen Übernahme des erhaltenen Ratschlags (100%- bzw. Vollgewichtung). Bei dieser Wahl entscheidet sich die herausragende Menge der Versuchsteilnehmer allerdings für die Beibehaltung der eigenen Schätzung und damit vollständige Diskontierung des Ratschlags.

Wie die ermittelte grafische Form des Zusammenhangs zwischen Ratschlagsdistanz und –gewichtung (siehe Abbildung 2) zu Stande kommt, lässt sich an Hand der Aufschlüsselung der Gewichtungsstrategien nach Ratschlagsdistanzen ebenfalls gut erkennen: Die sehr niedrige mittlere Gewichtung naher Ratschläge (bis 15% Abweichung) resultiert aus dem überwiegenden Anteil von Nullgewichtungen (zwischen 46% und 69% Häufigkeit) bei den betreffenden Abweichungen. Erst danach, das heißt ab Abweichungen von 20% und mehr nehmen sowohl Gewichtungen im Rahmen von 40% bis 60%, aber auch von 20% bis unter 40% zu, während zur selben Zeit der Anteil der Nullgewichtungen stetig abnimmt. Dadurch ergeben sich in diesen Distanzbereichen dann auch höhere AT-Mittelwerte. Ab Ratschlagsabweichungen von 65% nimmt dann insbesondere der Anteil der *Averaging*-Strategie wieder ab bei gleichzeitigem Anstieg der Häufigkeit der Anwendung von Nullgewichtungen. Auf den mittleren AT-Wert wirkt sich dies dann wieder vermindernd aus, wenngleich auch nicht so niedrige Werte erzielt werden wie bei sehr nahen Abweichungen, da einerseits der Anteil der Nullgewichtungen deutlich unter dem Niveau dessen naher Ratschläge bleibt, während gleichzeitig Gewichtungen zwischen 20% bis knapp 40% immernoch recht häufig bzw. häufiger als bei nahen Abweichungen vorkommen.

4.3.3 Urteilssicherheit

Jeweils im Anschluss an ihre Initial- und Finalschätzungen wurden die Judges aufgefordert, ihre Urteilssicherheit hinsichtlich dieser Schätzungen anzugeben. Pro Versuchsdurchgang wurde die Differenz aus Final- und Initialsicherheit gebildet. Eine positive Differenz würde bedeuten, dass der Judge während eines Trials Urteilssicherheit gewonnen hat, das heißt also nach Darbietung des Ratschlags noch mehr Vertrau-

en in sein finales Urteils besitzt; eine negative Differenz würde dementsprechend das Gegenteil implizieren, nämlich, dass die Urteiler nach der Darbietung des Ratschlags eher verunsichert werden. Analog zur Datenaufbereitung bei den AT-Maßen wurde zunächst pro Proband und Trial die Differenz aus Final- und Initialschätzung berechnet. Im Anschluss wurden diese Differenzen (im Folgenden auch vereinfacht als „Urteilssicherheit“ bezeichnet, obwohl es genau genommen „Veränderung der Urteilssicherheit“ heißen müsste) hinsichtlich der Ratschlagsabweichungen, bei denen sie auftraten, angeordnet.

Die Varianzanalyse mit Messwiederholung, in der wiederum die Ratschlagsabweichung als 15-stufige unabhängige Variable fungierte, ergab, dass diese auch einen signifikanten Effekt auf die Urteilssicherheit ausübt, $F(4.97, 129.27) = 2.99$ $p = .014$, $\eta^2 = .10$. Des Weiteren erwies sich der postulierte lineare Kontrast als signifikant, $F(1, 26) = 4.51$, $p = .043$, $\eta^2 = .15$. In Einklang mit Hypothese 2 erlangten die Versuchsteilnehmer nach dem Erhalt eines Ratschlags im Durchschnitt einen Sicherheitszuwachs ($M_{\text{gesamt}} = .41$, $SD = .70$), jedoch war dieser Effekt für nahe Ratschläge stärker ausgeprägt als für weit entfernte. Das heißt, mit zunehmender Abweichung der Ratschläge nahm die Urteilssicherheit ab. Entgegen der Hypothese wurde allerdings auch der quadratische Kontrast signifikant, $F(1, 26) = 29.87$, $p < .001$, $\eta^2 = .54$ (siehe auch Trendlinie in Abbildung 5). Dementsprechend nimmt die Urteilssicherheit ab einem bestimmten Punkt trotz hoher Abweichungen der Ratschläge augenscheinlich wieder zu.

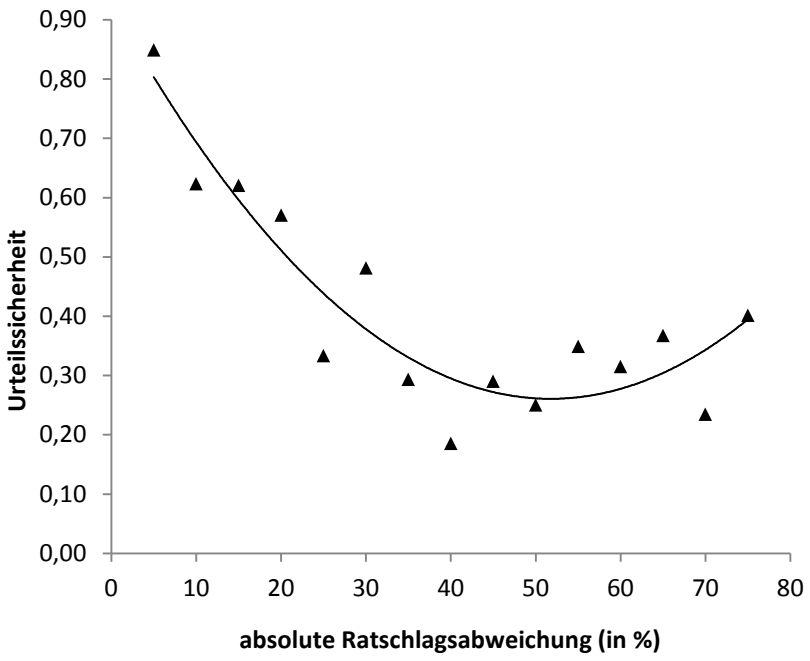

Abbildung 5

grafische Beziehung zwischen Ratschlagsdistanz (in %) und mittlerer Veränderung der Urteilssicherheit in Experiment 1

Bei Betrachtung der gegenläufigen quadratischen Trends für den Zusammenhang von Ratschlagsdistanz mit Ratschlagsgewichtung und Urteilssicherheit, könnte man zu dem Schluss gelangen, dass die beiden AVn sich stets konträr verhalten, sprich, dass also beispielsweise eine abnehmende Gewichtung überwiegend mit einem Anstieg der Urteilssicherheit einhergeht und umgekehrt. Die Korrelation der jeweiligen Mittelwerte beider AVn pro Abweichung ist hoch negativ mit $r(15) = -.82$, $p < .001$. Wenn die Korrelationen zwischen Ratschlagsgewichtung und Urteilssicherheit ungeachtet der ihnen vorausgehenden Abweichungen für jede Versuchsperson einzeln berechnet werden ergibt sich im Mittel eine negative Korrelation von $r = -.09$. Außerdem zeigt sich, dass die Anzahl der negativen Zusammenhänge gegenüber positiven zwischen beiden Größen überwiegt, χ^2 (1, N = 26) = 3.85, $p = .05$. Mit anderen Worten scheint eine Vorhersage der einen Variablen aus der anderen möglich.

4.3.4 Explorative Datenanalyse: Schätzungsakkuratheit der Judges

Im Anschluss an die Hauptanalysen erfolgte eine explorative Analyse der Schätzungsakkuratheit der Judges und deren Zusammenhang mit den gefundenen Distanz-

effekten bei der Ratschlagsgewichtung. In Anlehnung an Yaniv (2004b) sollte hierbei insbesondere überprüft werden, ob sich für unterschiedlich gute Versuchspersonen der gefundene Distanzeffekt bei der Ratschlagsgewichtung auch unterschiedlich konsistent oder in unterschiedlicher Form zeigt. Als objektives Maß für die Urteilsakkuratheit wurden jeweils getrennt für die Initial- und die Finalschätzungen die tatsächliche Abweichung und die betragsmäßige Abweichung vom wahren Wert berechnet.[3] Sowohl für die tatsächliche als auch für die betragsmäßige prozentuale Abweichung wurden im Anschluss Mittelwerte gebildet. Es resultierten so pro Person und jeweils für Initial- und Finalschätzung der *mean Percentage Error (MPE)* bzw. der *mean absolute Percentage Error (MAPE)*. Die tatsächliche prozentuale Abweichung ermöglicht an Hand ihres Vorzeichens im Gegensatz zum betragsmäßigen prozentualen Abweichung Aussagen darüber, in welchen einzelnen Trials der Proband mit seiner Schätzung den wahren Wert unter- oder überschätzt hat. Werden diese Werte dann zum Beispiel über alle Initialschätzungen gemittelt, lässt sich so möglicherweise eine generelle Tendenz des Probanden zur Über- oder Unterschätzung ablesen. Es ist aber auch denkbar und möglich, dass sich Über- und Unterschätzungen bei den Initialschätzungen eines Probanden über den gesamten Versuch hinweg intraindividuell relativ ausgleichen, sodass im Mittel für ihn dann eine nur sehr geringe bis gar keine Abweichung resultiert. Diese im Anschluss als hohe mittlere Schätzungsakkuratheit des Judges zu interpretieren, wäre dann allerdings inhaltlich falsch. Daraus wird ersichtlich, dass die Betrachtung der betragsmäßigen Abweichung pro Versuchsdurchgang und in Folge auch deren Mittelwert unter Umständen aussagekräftiger ist, nämlich dann, wenn es um den Umfang der Abweichung der Schätzungen vom wahren Wert geht. Der MAPE ist deshalb das bewährte Maß, wenn es um die Beurteilung der Qualität von quantitativen Urteilen geht (vgl. bspw. Sniezek & Henry, 1989, 1990). Im Folgenden konzentriere ich mich dabei vor allem auf die MPE- bzw. MAPE-Werte für die *Initial*schätzungen der Probanden, weil diese Anfangsschätzungen der Urteiler in Eigenleistung erfolgen und damit den besten Indikator für die judge-eigene Kompetenz darstellen. Eine Analyse

[3] Formel für die tatsächliche Abweichung vom wahren Wert: $\frac{wahrer\ Wert - Schätzung}{wahrer\ Wert}$. Die betragsmäßige Abweichung vom wahren Wert errechnet sich demgemäß: $\frac{|wahrer\ Wert - Schätzung|}{wahrer\ Wert}$.

der MPE- und MAPE-Werte für die Finalschätzung erscheint wenig aussagekräftig, da es sich bei den dargebotenen Ratschlägen, wie bereits beschrieben, jeweils um Linear-transformationen der Initialschätzungen der Probanden handelt. Demgemäß dürfte die Qualität der dargebotenen Ratschläge im Mittel nicht wesentlich von der Qualität der Initialschätzungen abweichen. Daraus folgt wiederum, dass sich die Qualität der Final-schätzungen unter Einbezug der Ratschläge weder systematisch verbessern noch ver-schlechtern sollte. Empirisch untermauert wird dies durch die Tatsache, dass MPE und MAPE von Initial- und Finalschätzungen in Experiment 1 jeweils im Mittel nahezu iden-tisch sind (MPE: $M_{\text{initial,}}$ = 0.21, *SD* = 0.86 und M_{final} = 0.21, *SD* = 0.87; MAPE: $M_{\text{initial,}}$= 0.72, *SD* = 0.67 und M_{final} = 0.72, *SD* = 0.70), t_{MPE} (26) = 0.02, *p* = .99, *d* = .05 bzw. t_{MAPE} (26) = -0.03, *p* = .98, *d* = .04.

Der über alle Probanden gemittelte MPE deutet mit einem Wert von 0.21 da-rauf hin, dass die Entfernungen zwischen europäischen Hauptstädten grundsätzlich eher überschätzt werden, nämlich um etwa 21% (M_{MPE} = 0.21, *SD* = .86). Allerdings ist diese Tendenz nicht signifikant verschieden von 0, *t*(26) = 1.30, *p* = .21, *d* = .25, sodass eine systematische Überschätzung nicht nachgewiesen werden kann.

Wie geplant wurde nun analysiert, ob der gefundene Distanzeffekt möglicher-weise wie auch bei Yaniv (2004b) durch die Schätzungsakkuratheit der Probanden be-einflusst wird. Zu diesem Zwecke wurden die Probanden an Hand ihrer Leistungen (MAPE) bei den Initialschätzungen per Mediansplit zunächst in zwei Gruppen hinsicht-lich ihrer Schätzungsakkuratheit (hoch vs. niedrig) aufgeteilt.[4] Im Anschluss wurden dann die Haupt-ANOVAs für die Ratschlagsgewichtung und die Urteilssicherheit jeweils um den Zwischensubjektfaktor *Schätzungsakkuratheit der Judges* (hoch vs. niedrig) ergänzt. Dabei zeigte sich ein signifikanter Haupteffekt dieser Akkuratheit, *F*(1, 24) = 7.34, *p* = .012, η^2= .23, wobei akkuratere Versuchspersonen erhaltene Ratschläge grundsätzlich weniger gewichteten (*M* = .22, *SD* = .22) als weniger akkurate Versuchs-personen (*M* = .42, *SD* = .23). Der signifikante Distanzeffekt, *F*(7.21, 173.03) = 6.21, *p* <

[4] Die Einteilung in zwei Probandengruppen bringt Informationseinbußen mit sich, wurde jedoch trotzdem angewendet, um eine maximale Vergleichbarkeit mit der Originalstudie von Yaniv (2004b) zu erreichen.

.001, η^2= .21, wurde nicht durch die Schätzungsakkuratheit der Judges moderiert, was durch eine nicht-signifikante Interaktion angezeigt wurde, $F(7.21, 173.03) = 1.60$, $p = .14$, η^2= .06. Der Distanzeffekt scheint demnach weitgehend unabhängig von der Schätzungskompetenz der Judges zu sein. In Abbildung 6 kann man das höhere Gewichtungsniveau bei weniger akkuraten Versuchspersonen im Gegensatz zu akkuraten Versuchspersonen deutlich erkennen. Es zeigt sich zudem, dass die besonders niedrige Gewichtung sehr naher Ratschläge (bis 15% Abweichung) gegenüber den verbleibenden Distanzen bei den weniger akkuraten Versuchspersonen ein wenig deutlicher ausfällt (.26 bei nahen gegenüber .46 bei den entfernteren Ratschlägen) als bei den besseren Versuchspersonen (.15 bei sehr nahen und .23 bei den entfernteren Ratschlägen); auf Grund der Nicht-Signifikanz der Interaktion soll dies aber nicht näher interpretiert werden.

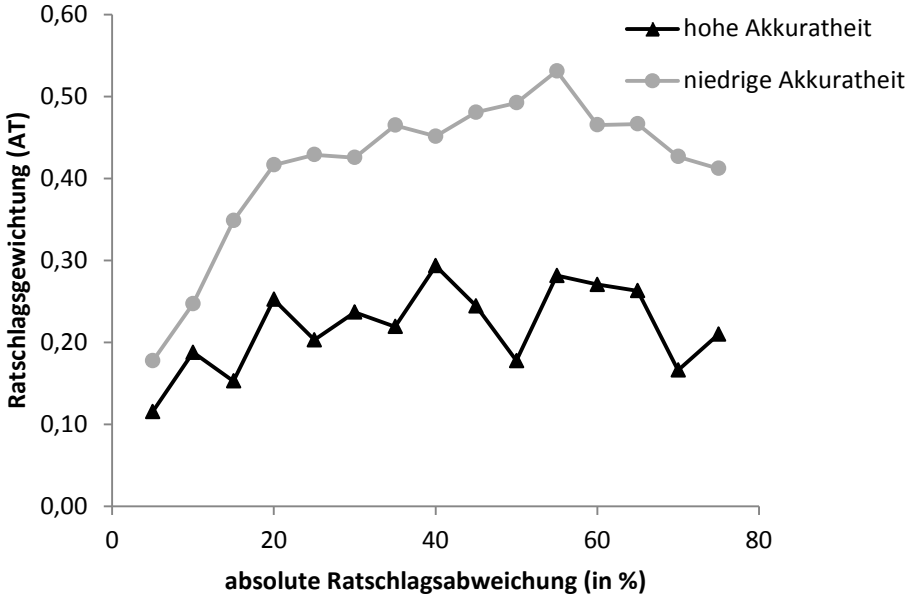

Abbildung 6

grafische Beziehung zwischen Ratschlagsdistanz (in %) und mittlerer Ratschlagsgewichtung in Abhängigkeit von der Schätzungsakkuratheit der Probanden in Experiment 1

Bei einer Analyse der Urteilssicherheit unter Berücksichtigung der Schätzungsakkuratheit der Judges zeigt sich für sie kein signifikanter Haupteffekt, $F(1, 24) = 0.64$, $p = .43$, η^2= .03, obwohl weniger akkurate Versuchspersonen im Laufe eines Versuchsdurchgangs rein deskriptiv deutlicher an Sicherheit dazu gewannen ($M = .51$, $SD = 1.08$)

als akkurate Versuchspersonen (*M* = .29, *SD* = .56). Die signifikante Abhängigkeit der Urteilssicherheit von der Ratschlagsdistanz, $F(5.41, 129.75) = 2.34$, $p = .04$, $\eta^2 = .09$, wurde nicht durch Schätzungsakkuratheit der Judges beeinflusst. Das heißt, die Interaktion zwischen Distanz und Schätzungsakkuratheit wurde nicht signifikant, $F(5.41, 129.75) = 1.38$, $p = .23$, $\eta^2 = .05$.

Zusammengefasst zeigen die Ergebnisse aus Experiment 1, dass nahe Ratschläge von den Probanden sehr wahrscheinlich als soziale Validierung ihrer eigenen Initialschätzungen wahrgenommen werden. Obwohl die Urteiler ihre Finalschätzungen diesen Ratschlägen nur kaum bis gar nicht angepasst haben, sind sie zur selben Zeit durch sie sicherer hinsichtlich ihrer eigenen (Initial)Schätzungen geworden. Mit anderen Worten wurden in Experiment 1 de facto auch nahe Ratschläge genutzt. Dies geschah jedoch nicht in Form einer Adjustierung der Finalschätzung, sondern vielmehr durch einen Zuwachs an Überzeugung bzw. Sicherheit, dass der Urteiler von vornherein richtig lag. Der informationale Wert naher Ratschläge spiegelt sich daher nicht in einer Urteilsanpassung, sondern in einer gesteigerten Urteilssicherheit wider. Gleichzeitig liefern damit die Ergebnisse aus Experiment 1 keine Unterstützung für Yanivs Annahme eines monoton fallenden Zusammenhangs zwischen Ratschlagsdistanz und -gewichtung. Stattdessen liefern sie Hinweise für das Zutreffen von Hypothese 1 in diesem Experiment: Die Beziehung zwischen Ratschlagsdistanz und -gewichtung scheint demnach kurvilinear zu sein. Diese kurvilineare Beziehung rührt vorwiegend daher, dass Ratschläge, die sehr nah an der Intitialschätzung des Ratgebers liegen sehr wenig gewichtet werden. Gleichzeitig führen genau diese Ratschläge zu einem sehr deutlichen Anstieg der Urteilssicherheit bei den Judges. Dies spricht dafür, dass die nahe Ratschläge als validierend (im Hinblick auf die zuvor abgegebene Initialschätzung) empfunden werden.

Bei der Analyse der vorkommenden Gewichtungen zeigte sich im Detail wie die Form des gefundenen Zusammenhangs zwischen Ratschlagsdistanz und -gewichtung entstehen konnte: Die Probanden zeigten in Abhängigkeit der Entfernung des Ratschlags unterschiedlich häufig insbesondere drei Gewichtungsstrategien; die Nullgewichtung, das *Averaging* (Gewichtungen zwischen 40% und 60%) sowie Gewichtungen

zwischen 20% und knapp 40%. Sehr nahe Ratschläge wurden überwiegend zu 0% gewichtet, was die niedrigen AT-Mittelwerte im Bereich dieser Abweichungen erklärt. Mit zunehmenden Distanzen nahm diese Gewichtungstrategie jedoch fast durchgängig ab, während die anderen beiden genannten Strategien häufiger gezeigt und die AT-Mittelwerte dadurch insgesamt größer wurden. Bei den höchsten Abweichungen verhielt sich diese Entwicklung dann wieder rückläufig und niedrigere AT-Mittelwerte resultierten, wenngleich diese nicht so niedrig ausfielen wie bei nahen Ratschlägen, da der Anteil von Nullgewichtungen hier immernoch deutlich niedriger ausfiel.

Zusätzlich zeigen die Ergebnisse der explorativen Analysen, dass die gefundenen Distanzeffekte relativ unabhängig von der judge-eigenen Expertise sind. Mit anderen Worten zeigen also akkurate und weniger akkurate Versuchspersonen weitestgehend die gleichen Zusammenhänge zwischen Ratschlagsdistanz und den beiden gemessenen AVn. Nur das grundlegende Ausprägungsniveau unterschied sich für die unterschiedlich gut schätzenden Gruppen, im Falle der Urteilssicherheit jedoch nur rein deskriptiv. So scheinen besser schätzende Probanden Ratschläge insgesamt eher in geringerem Umfang zu nutzen, als weniger kompetent schätzende Personen. Die niedrigere Nutzung durch akkurater schätzende Probanden manifestiert sich dabei sowohl in einer deutlich geringeren Gewichtung von Ratschlägen, als auch in tendenziell weniger Sicherheitszuwachs durch derlei Hilfestellungen.

5. Experiment 2

5.1 Zielsetzung und Hypothesen

Experiment 2 wurde mit dem Ziel durchgeführt, die Übertragbarkeit der Ergebnisse aus Experiment 1 auf einen anderen Urteilsaufgabentyp zu überprüfen. Bei Experiment 2 handelt es sich demnach um ein Generalisierungsexperiment. Es liegen daher dieselben zwei Hypothesen wie in Experiment 1 zu Grunde.

Hypothese 1: Die Ratschlagsgewichtung ist eine quadratische (umgekehrt U-förmige) Funktion der Distanz zwischen Initialschätzung und Ratschlag, das heißt Ratschläge werden weniger stark gewichtet, wenn sie entweder sehr nah an oder sehr

weit von der Initialschätzung des Judges entfernt sind, im Vergleich zu Ratschlägen, deren Entfernung zur Initialschätzung moderat ausfällt.

Hypothese 2: Je näher Ratschläge an der Initialschätzung liegen, desto stärker wird die Urteilssicherheit zwischen Initial- und Finalurteil ansteigen.

5.2 Methode

5.2.1. Stichprobe und Design

Insgesamt nahmen 34 Versuchspersonen an Experiment 2 teil. Drei von ihnen mussten ausgeschlossen werden, weil entweder –vermutlich durch eine Instabilität des Computer-Programms– das Experiment nicht vollständig durchgeführt werden konnte (n = 2) oder aber der Proband die Manipulation dahingehend in Zweifel zog, dass der Ratgeber tatsächlich existiert bzw. es sich demnach bei seinen Ratschlägen um echte Schätzungen handelt (n = 1).[5] Daher gingen die Daten von den verbleibenden 31 Probanden in die statistischen Analysen ein. Die Analysestichprobe bestand aus 18 (58%) weiblichen und 13 männlichen Probanden. Die Teilnahme wurde entweder durch Versuchspersonenstunden oder durch monetäre Vergütung (in Höhe von 5-6€) entgolten. Erneut wurden zur Motivationssteigerung Kino-Gutscheine unter den akkuratesten fünf verlost.

Mit dem Ziel der Replikation der Befunde aus Experiment 1 wurde auch in Experiment 2 ein einfaktorielles Design mit dem 15-stufigen Innersubjektfaktor *Ratschlagsdistanz* (5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%) angewendet.

5.2.2 Ablauf

Bei Experiment 2 handelt es um eine Replikation von Experiment 1, mit einer grundlegenden Modifikation, nämlich dem verwendeten Schätzaufgabentyp. Anstelle von Luftlinienentfernungen wurden die Probanden nun dazu aufgefordert, den Nähr-

[5] Zitat des Probanden: „[Es ging in dem Versuch] um den Einfluss von Christian, der nicht wirklich existierte und ziemlich dämliche Angaben gemacht hat.“

wert (Kaloriengehalt) bestimmter Lebensmittel zu schätzen. Der drei-phasige Ablauf der Versuchsdurchgänge wurde beibehalten: Zuerst wurden die Probanden dazu aufgefordert, den Kaloriengehalt eines bestimmten Lebensmittels zu schätzen (z.B. Toastbrot, Erbsen, gekochtes Ei, jeweils 100g) und ihre Urteilssicherheit hinsichtlich dieser Initialschätzung abzugeben. Im Anschluss an diese erste Phase erhielten die Probanden Ratschläge, welche wiederum hinsichtlich ihrer Abweichung von der Initialschätzung variierten. Die Variation umfasste erneut einen Bereich von ±5% bis ±75% und erfolgte in 5%-Schritten, die diesmal ad-hoc zufällig von Computer ausgewählt wurden. In der dritten Phase eines jeden Versuchsdurchgangs wurden die Probanden dann gebeten, eine Finalschätzung hinsichtlich des Kaloriengehalts des erwähnten Lebensmittels sowie ihre subjektive Urteilssicherheit hinsichtlich dieser Schätzung abzugeben. Jede positive und negative Ratschlagsabweichung wurde jeweils zweimal angewendet, wodurch das Experiment aus 60 Versuchsdurchgängen pro Teilnehmer bestand.

Wie schon in Experiment 1 wurde den Versuchsteilnehmern erklärt, dass sie die Schätzungen eines Probanden aus einem früheren Experiment mit denselben Aufgaben als Ratschläge erhalten würden. Dieser Person –die laut Instruktion dasselbe Geschlecht wie die Versuchsteilnehmer besaß– wurde eine relativ hohe aufgabenbezogene Kompetenz zugesprochen, welche durch ihren siebten Rang unter den Leistungen der 100 früheren Teilnehmer verdeutlicht wurde.

5.3 Ergebnisse und Diskussion

5.3.1 Analyse möglicher Störvariablen

Gemäß der in Experiment 1 aufgeführten Überlegungen umfasste auch in Experiment 2 der Gültigkeitsbereich des AT die Werte zwischen -1.5 und +1.5. Weiterhin wurden auch die Versuchsdurchgänge ausgeschlossen, in denen die Initialschätzung entweder so hoch oder niedrig ausfiel, dass als Ratschlag eine Zufallszahl dargeboten wurde. Diesen beiden Ausschlusskriterien folgend mussten insgesamt 330 von 1860 gültigen Versuchsdurchgängen ausgeschlossen werden. Mit einer dementsprechenden Ausschlussrate von knapp 18% wurde ein vergleichsweise hoher Wert erreicht. Bei genauerer Betrachtung wird der überwiegende Teil an Ausschlüssen dadurch produ-

ziert, dass statt der üblichen Lineartransformation eine Zufallszahl als Ratschlag dargeboten worden war. Daran zeigt sich, dass der Bereich der gültigen Werte für die Ratschläge möglicherweise zu eng gewählt wurde.[6] Die verbleibenden AT-Werte gingen in die Hauptanalyse mit ein und wurden auch zur Bestimmung eines mittleren AT-Wertes genutzt. Bei Korrelation dieses Wertes mit dem Alter der Probanden zeigte sich erneut kein systematischer Zusammenhang zwischen diesen beiden Variablen, $r(31) = -.05$, $p = .78$. Auch die Korrelation zwischen Alter und Urteilssicherheit fiel insignifikant aus, $r(31) = .08$, $p = .68$. Um zu überprüfen, ob das Geschlecht der Probanden möglicherweise einen systematischen Einfluss auf die Ratschlagsgewichtung und/oder die Urteilssicherheit ausübt, wurde jeweils wiederum eine ANOVA mit dem Geschlecht als Zwischen- und der Ratschlagsdistanz als Innersubjektfaktor berechnet. Hier zeigte sich weder ein signifikanter Effekt des Geschlechts auf die Ratschlagsgewichtung, $F(1, 27) = 0.00$, $p = .98$, $\eta^2 < .001$, noch eine signifikante Interaktion mit der Ratschlagsdistanz, $F(5.95, 160.62) = 0.83$, $p = .54$, $\eta^2 = .03$. Auch in Bezug auf die AV Urteilssicherheit erlangte der Haupteffekt Geschlecht keine Signifikanz, $F(1, 27) = 0.47$, $p = .501$, $\eta^2 = .02$. Zusätzlich gab es auch hier keine signifikante Interaktion mit der Ratschlagsdistanz, $F(5.797, 156.53) = 0.93$, $p = .47$, $\eta^2 = .03$. Ein systematischer Einfluss der Störvariablen auf die untersuchten AVn kann somit ausgeschlossen werden.

5.3.2 Ratschlagsgewichtung

Erneut wurde die Analyse der Ratschlagsgewichtung durch die Berechnung des AT-Maßes pro Person und Versuchsdurchgang vorbereitet. Im Anschluss wurden die ermittelten AT-Koeffizienten im Hinblick auf die jeweiligen Ratschlagsabweichungen, denen sie folgten, umgeordnet. Der Messwiederholungsfaktor Ratschlagsdistanz war erneut 15-stufig, und für ihn ergab sich ein signifikanter Haupteffekt, $F(6.01, 168.22) = 6.25$, $p < .001$, $\eta^2 = .18$, das heißt dass sich die Ratschlagsgewichtung in Abhängigkeit von der Abweichung der Ratschläge veränderte. Auch in Experiment 2 wurden die gefundenen Kontraste betrachtet, um eine Aussage treffen zu können wie die Beziehung

[6] Bei den Kalorienschätzungen wurde die Ratschlagsuntergrenze auf 10 und die Obergrenze auf 2000 gesetzt. Die im Falle von berechneten Ratschlagswerten außerhalb dieser Grenzwerte dargebotenen Zufallszahlen lagen im Bereich von 20 bis 900.

zwischen Ratschlagsdistanz und –gewichtung geartet ist. Es zeigten sich auch hier sowohl ein signifikant linearer, $F(1, 28) = 13.80$, $p = .001$, $\eta^2 = .33$, als auch ein signifikant quadratischer Kontrast, $F(1, 28) = 30.489$, $p < .001$, $\eta^2 = .52$ (siehe auch Abbildung 7).

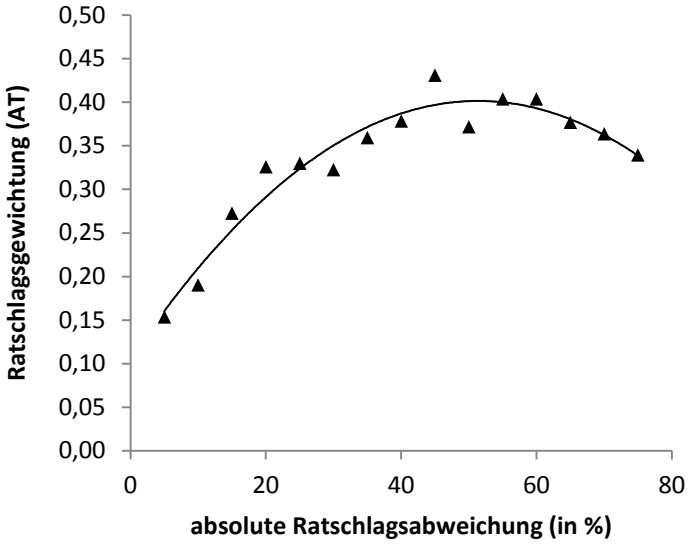

Abbildung 7

grafische Beziehung zwischen Ratschlagsdistanz (in %) und mittlerer Ratschlagsgewichtung in Experiment 2

Wie erwartet übte also die Ratschlagsdistanz einen großen Einfluss auf das Ausmaß der Ratschlagsgewichtung aus. Die Art dieses Einflusses ist dabei quadratisch und umgekehrt U-förmig, was sich mit der Vorhersage aus Hypothese 1 deckt, welche daher als bestätigt angesehen werden kann. Zusätzlich zur statistischen Signifikanz des quadratischen Kontrasts zeigt auch die Abbildung 7 die charakteristische Form dieser Beziehung. Wie an Hand dieser Abbildung auch zu erkennen ist, fällt das mittlere Gewichtungsniveau bei entfernten Ratschlägen, trotz umgekehrter U-Form, nicht so niedrig aus wie bei sehr nahen Ratschlägen, weswegen auch in Experiment 2 der lineare Kontrast Signifikanz erlangte.

Zum besseren Verständnis des Zustandekommens der mittleren Ratschlagsgewichtungen pro Ratschlagsdistanz wurde im Anschluss erneut der Verteilung der Gewichtungen analysiert. Die Vorgehensweise entsprach der von Experiment 1. Zunächst wurden die ermittelten AT-Werte pro Person und Trial demnach kategorisiert und im

Anschluss die Auftretenshäufigkeit der einzelnen Kategorien pro absoluter Ratschlagsabweichung ausgezählt. Bei Betrachtung der Gewichtungsverteilung ohne Berücksichtigung der Ratschlagsabweichung zeigte sich auch in Experiment 2 eine zwei-gipflige Verteilung, mit den höchsten Raten bei der Null-Gewichtung und der (ungefähr) 50%-Gewichtung. Die dritthäufigste Kategorie beinhaltete auch hier Gewichtungswerte im Bereich von 20% bis unter 40% (Abbildung 8).

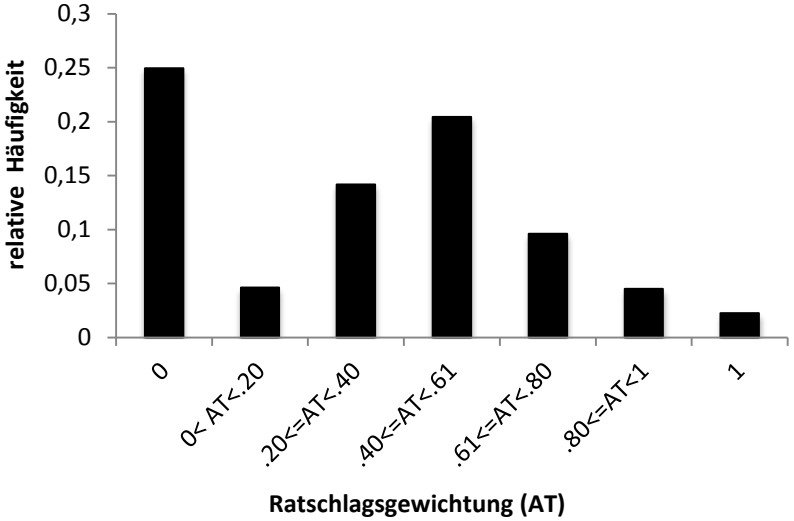

Abbildung 8

Verteilung der Ratschlagsgewichtungen in Experiment 2

Wie in Abbildung 9 zu erkennen ist, ähneln die Ergebnisse hinsichtlich der Gewichtungsverteilungen pro Ratschlagsdistanz aus Experiment 2 denen von Experiment 1 deutlich. So wurden bei niedrigen Distanzen Ratschläge in der Mehrzahl vollständig diskontiert, allenfalls noch mit der Initialschätzung zu einem Finalurteil gemittelt, was die niedrige mittlere Gewichtung von Ratschlägen mit diesen Distanzen erklärt. Mit zunehmender Ratschlagsdistanz änderte sich die bevorzugte Gewichtungsstrategie zu einer mehrheitlichen Mittelwert-Bildung aus der eigenen Schätzung und den erhaltenen Ratschlägen oder aber teilweise auch dahingehend, dass Ratschlage im Umfang von 20% bis knapp 40% gewichtet wurden. Bei großen Distanzen nahmen die geringfügigen Gewichtungen (bis knapp 20% und zwischen 20% und unter 40%) zu, während die *Averaging*-Strategie weniger gezeigt wurde. Die Folge sind abnehmende AT-Werte bei diesen großen Distanzen, die allerdings wiederum nicht so gering ausfallen wie bei

sehr nahen Distanzen, da hier einerseits Nullgewichtungen deutlich weniger und diese geringfügigen Gewichtungen häufiger vorkommen.

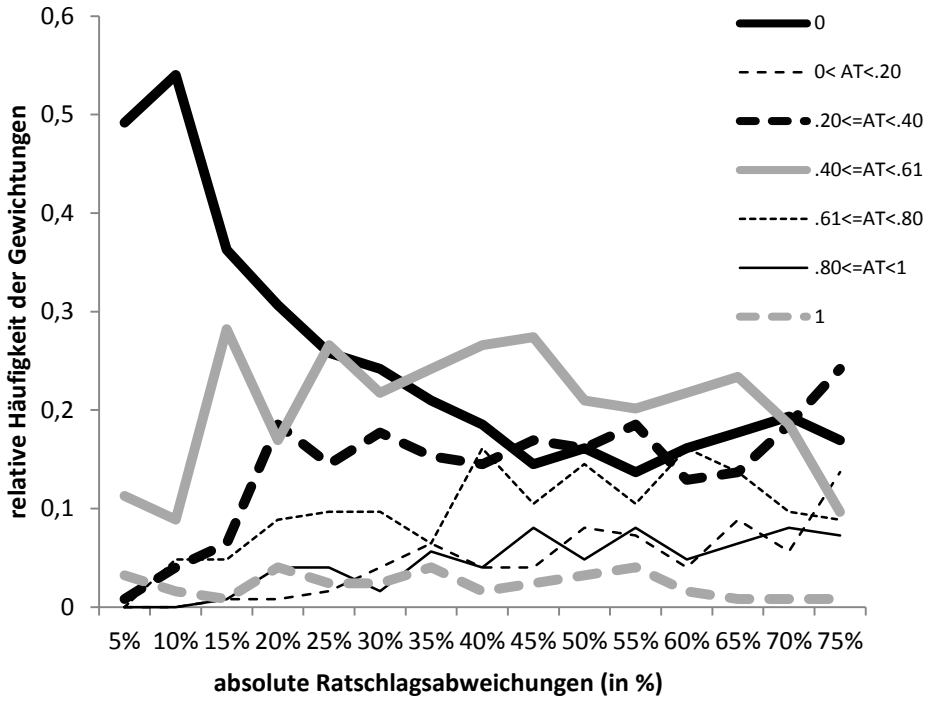

Abbildung 9

relative Häufigkeiten der Ratschlagsgewichtungen in Abhängigkeit der absoluten Ratschlagsentfernung in Experiment 2

5.3.3 Urteilssicherheit

Die Veränderung der Urteilssicherheit von Initial- zu Finalschätzung wurde zunächst für jeden Versuchsdurchgang und jeden Probanden berechnet. Im Anschluss wurden diese Differenzen den mit ihnen verbundenen absoluten Ratschlagsabweichungen zugeordnet. Allgemein nahm die Urteilssicherheit der Versuchsteilnehmer, zwischen ihrer Initialschätzung und dem dargebotenen Ratschlag, im Laufe eines Versuchsdurchgangs zu (M = .45, SD = .47). Wie schon zuvor bei der Ratschlagsgewichtung, zeigte die rm-ANOVA allerdings auch für die abhängige Variable der Urteilssicherheit einen signifikanten Haupteffekt für die Ratschlagsdistanz, $F(5.88, 164.59) = 9.83$, $p < .001$, $\eta^2 = .26$. Gemäß Hypothese 2 zur Art der Beziehung zwischen Ratschlagsdistanz und Urteilssicherheit erlangte der lineare Kontrast Signifikanz, $F(1, 28) = 45.82$, $p < .001$, $\eta^2 = .62$. Wie bereits in Experiment 1 erwies sich unerwarteterweise

auch der quadratische Kontrast als signifikant, $F(1, 28) = 8.54$, $p = .007$, $\eta^2 = .23$, was sich vermutlich durch den erneuten Anstieg der mittleren Urteilssicherheiten bei einigen weit entfernten Ratschlagsdistanzen ergeben hatte (siehe Abbildung 10).

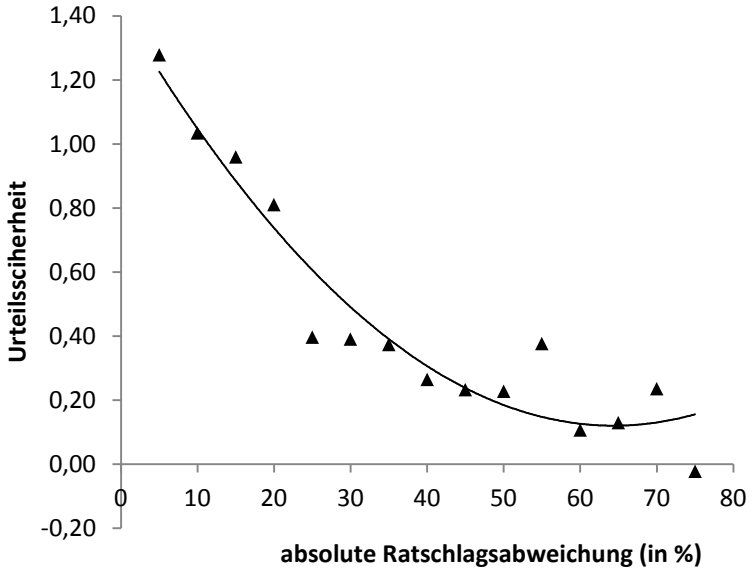

Abbildung 10

grafische Beziehung zwischen Ratschlagsdistanz (in %) und mittlerer Veränderung der Urteilssicherheit in Experiment 2

Die Korrelation der Mittelwerte von Ratschlagsgewichtung und Urteilssicherheit in Abhängigkeit der verwendeten Ratschlagsdistanzen zeigte auch in Experiment einen hohen negativen Zusammenhang zwischen den beiden AVn, $r(15) = -.87$, $p < .001$. Das heißt durch Kenntnis des Verlaufs einer Variable in Abhängigkeit der Ratschlagsdistanz lässt sich die Ausprägung der anderen ziemlich genau vorhersagen. Steigt die Ratschlagsgewichtung von einer zur nächsten Abweichung an, kann davon ausgegangen werden, dass die mittlere Urteilssicherheit zwischen denselben beiden Abweichungsstufen sinkt. Werden die Korrelationen zwischen Ratschlagsgewichtung und Urteilssicherheit individuell pro Versuchsperson berechnet ergibt sich ein ähnlicher mittlerer Zusammenhang wie in Experiment 1 von $r = -.07$. Dieses Mal liegt allerdings kein signifikant häufigeres Vorkommen von negativen Korrelationen gegenüber positiven vor, χ^2 $(1, N = 29) = 0.86$, $p = .35$. Auf individueller Ebene kommt die Gegensätzlichkeit von Ratschlagsgewichtung und Urteilssicherheit also nicht systematisch häufiger vor, als dass beide Variablen in die gleiche Richtung kovariieren. Eine Erklärung dafür, dass sich

bei Mittelung der Ratschlagsgewichtung und der Urteilssicherheit dann ein so deutlich gegenläufiger Trend ergibt könnte sein, dass Abnahmen in der Urteilssicherheit bei gesteigerter Ratschlagsgewichtung numerisch größer ausfallen als entsprechende Anstiege und das somit bei Mittelung über alle Probanden hinweg erstere stärker ins Gewicht fallen. Eine Korrelation zeigt ja letztendlich nur die Stärke der gemeinsamen Variation zweier Variablen an, nicht aber in welchem Ausmaß sie sich jeweils verändern.

5.3.4 Explorative Datenanalyse: Schätzungsakkuratheit der Judges

Auch in Experiment 2 wurde explorativ untersucht, welche Zusammenhänge zwischen Akkuratheit der Schätzungen der Judges und verschiedenen anderen Variablen existieren. Ergänzend gab es in Experiment 2 auch ein subjektives Maß für die Einschätzung der Schätzungsakkuratheit der Judges. Dafür wurden diese nach Beendigung aller Versuchsdurchgänge gefragt, welchen Rang unter 100 (angeblichen) Versuchsteilnehmern sie ihrer eigenen Schätzleistung geben würden. Rang 1 würde demgemäß widerspiegeln, dass der Judge seine Schätzungen subjektiv für die akkuratesten aller Versuchsteilnehmer hält, wohingegen ein selbsterteilter Rang 100 die Annahme der inakkuratesten Schätzleistung bedeutete. Als objektives Maß für die Urteilsakkuratheit wurden jeweils getrennt für die Initial- und die Finalschätzungen wieder die tatsächliche und die betragsmäßige Abweichung vom wahren Wert berechnet und im Anschluss Mittelwerte gebildet.

Im Hinblick auf die Schätzungen des Kaloriengehalts von Lebensmitteln neigten die Probanden aus Experiment 2 im Mittel zu einer systematischen Überschätzung um knapp 71% (M_{MPE} = .706, SD = 1.02), $t(30) = 3.85$, $p = .001$, $d = .69$. Bei der subjektiven Beurteilung der Qualität der eigenen Schätzurteile wurde am häufigsten die Rangplätze 20 und 25 angegeben, zusammen insgesamt 11 Mal (36%). Dies spiegelt sich auch im Mittelwert der subjektiv vermuteten Kompetenz wider, M = 26.68, SD = 18.08. In wieweit tatsächliche und vermutete Akkuratheit der Schätzungen übereinstimmten, wurde mittels der Korrelation zwischen der Einschätzung des eigenen Rangplatzes und dem MAPE-Wert bestimmt. Da hohe und niedrige Werte bei Rangplätzen und den MAPE-Werten jeweils dieselbe Merkmalsausprägung kennzeichnen (niedrige Zahlenwerte = hohe Akkuratheit; hohe Werte = niedrige Akkuratheit bzw. hohe Inakkurat-

heit), müsste demnach bei einer realistischen Einschätzung der eigenen Kompetenz eine hohe (positive) Korrelation zwischen beiden Werten bestehen. Dies ist jedoch nicht der Fall; genau genommen stehen subjektive und tatsächliche Akkuratheit der Schätzungen in nahezu keinerlei statistischen Zusammenhang, $r(31) = .09$, $p = .62$. Weiterhin zeigte sich, dass die subjektiv vermutete Akkuratheit weder einen signifikanten Zusammenhang mit der Ratschlagsgewichtung, $r(31) = -.05$, $p = .78$, noch mit der Urteilssicherheit, $r(31) = -.16$, $p = .39$, aufweist. Interessanterweise zeigt sich jedoch zum gleichen Zeitpunkt ein signifikanter Zusammenhang zwischen der objektiven Akkuratheit der Schätzungen und der mittleren Ratschlagsgewichtung, $r(31) = .46$, $p = .01$. Das heißt, je weniger akkurat (größerer MAPE-Wert) die Initialschätzungen der Probanden ausfielen, desto stärker gewichteten sie im Mittel die erhaltenen Ratschläge, das heißt also, dass weniger kompetente bzw. weniger akkurate Versuchspersonen den dargebotenen Ratschlag im Durchschnitt stärker Folge leisten als akkurater schätzende Personen. Angesichts der Tatsache, dass Probanden anscheinend nicht dazu in der Lage sind, ihre eigene Kompetenz bzw. Akkuratheit gut einzuschätzen, ist dieses Ergebnis ein wenig überraschend. Ebenso zeichnet sich ab, dass mit zunehmender Inakkuratheit der Initialschätzungen der Judges auch der Gewinn an Urteilssicherheit im Laufe eines Trials zunimmt, $r(31) = .43$, $p = .02$.

Auch in Experiment 2 wurden die Probanden an Hand der Akkuratheit ihrer Initialschätzungen per Mediansplit in zwei Gruppen (akkurate vs. weniger akkurate Versuchspersonen) eingeteilt. Im Hinblick auf die AV Ratschlagsgewichtung zeigte sich erneut, dass die Schätzungsakkuratheit die Höhe der Ratschlagsgewichtung systematisch beeinflusst, $F(1, 27) = 4.79$, $p = .037$, $\eta^2 = .15$. Im Mittel gewichteten Personen mit einer hohen Schätzungsakkuratheit die dargeboten Ratschläge wiederum weniger stark ($M = .28$, $SD = .21$), als Personen mit einer niedrigen Schätzungsakkuratheit bezüglich ihrer Initialschätzungen ($M = .41$, $SD = .23$). Weiterhin zeigte sich, dass der signifikante Distanzeffekt, $F(6.38, 172.12) = 7.02$, $p < .001$, $\eta^2 = .21$, durch die Schätzungsakkuratheit der Judges moderiert wurde, $F(6.38, 172.12) = 2.16$, $p = .045$, $\eta^2 = .07$. Bei separater Analyse der zwei Akkuratheits-Gruppen zeigte sich, dass der Distanzeffekt nur bei den weniger akkuraten Versuchspersonen Signifikanz erlangte, $F(5.41, 64.91) = 7.14$, $p < .001$, $\eta^2 = .37$, nicht jedoch bei den akkuraten Versuchspersonen, $F(4.60,$

69.06) = 1.82, p = .13, η^2= .11. In der Bedingung der weniger akkuraten Versuchspersonen zeigte sich zudem, dass der Distanzeffekt sowohl durch einen linearen, $F(1, 12)$ = 16.07, p = .002, η^2= .57, als auch durch einen quadratischen Kontrast, $F(1, 12)$ = 32.82, p < .001, η^2= .73, erklärt werden konnte. Bei Betrachtung der Abbildung 11 zeigt sich, dass sehr nahe Ratschläge (5% und 10% Abweichung) von den weniger akkuraten Schätzern deutlich weniger gewichtet wurden (im Mittel zu .15) als die Ratschläge mit den verbleibenden höheren Abweichungen (.45). Bei den Probanden mit akkurateren Initialschätzungen verläuft die Ratschlagsgewichtung über die verschiedenen Abweichungen hinweg dagegen deutlich ähnlicher (im Mittel .19 bei sehr nahen und .29 bei den restlichen Abweichungen). Akkuratere Versuchspersonen ließen sich also insgesamt nicht so stark durch Ratschläge beeinflussen wie weniger akkurat schätzende Personen. Auch scheint die Ratschlagsdistanz für erstere eine geringfügigere Rolle zu spielen, da sie kaum Einfluss auf die Ratschlagsgewichtung nimmt.

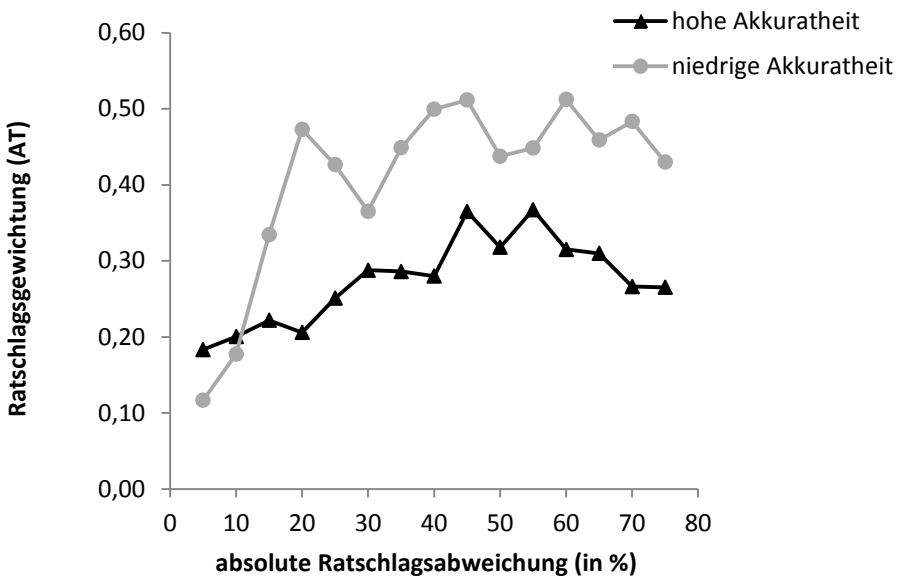

Abbildung 11

grafische Beziehung zwischen Ratschlagsdistanz (in %) und mittlerer Ratschlagsgewichtung in Abhängigkeit von der Schätzungsakkuratheit der Probanden in Experiment 2

Handelt es sich bei der mit der ANOVA analysierten AV um die Urteilssicherheit, ergibt sich im Gegensatz zu Experiment 1 ein signifikanter Haupteffekt für den Zwischensubjektfaktor Schätzungsakkuratheit, $F(1, 27)$ = 4.25, p = .049, η^2= .14. Im Detail

zeigte sich, dass die Probanden mit akkurateren Anfangsschätzungen im Laufe der Versuchsdurchgänge im Mittel weniger stark an Urteilssicherheit gewannen (M = .30, SD = .83) als Probanden mit weniger akkuraten Anfangsschätzungen (M = .64, SD = .65). Der Distanzeffekt für die Urteilssicherheit, $F(6.03, 162.896) = 9.42$, $p < .001$, $\eta^2 = .26$, wird dabei allerdings nicht systematisch von der Schätzungsakkuratheit der Judges beeinflusst, das heißt, es ergab sich keine signifikante Interaktion zwischen der Ratschlagsdistanz und der Akkuratheit, $F(6.03, 162.896) = 1.32$, $p = .25$, $\eta^2 = .05$.

Im Allgemeinen zeigen die Ergebnisse von Experiment 2, dass die gefundenen Beziehungen zwischen Ratschlagsabweichung und –gewichtung sowie Urteilssicherheit nicht auf einen speziellen Aufgabeninhalt beschränkt sind. Unabhängig davon, ob Luftlinienentfernungen oder Kaloriengehalt von Lebensmitteln geschätzt werden müssen, werden nahe Ratschläge deutlich weniger stark gewichtet als moderat oder weit entfernte. Zu den bevorzugtesten Gewichtungsstrategien zählten dabei wieder die vollständige Diskontierung des Ratschlags und somit Beibehaltung der eigenen Initialschätzung als Finalurteil sowie Bildung eines Mittelwerts aus Initialschätzung und Ratschlag zur Generierung des Endurteils. In Abhängigkeit der verschiedenen Ratschlagsdistanzen variierten die Häufigkeiten dieser Strategien, sodass sich die mittlere Gewichtung von Ratschlägen zwischen den einzelnen Ratschlagsdistanzen unterschied und die umgekehrte U-Form des Zusammenhangs zwischen Ratschlagsdistanz und -gewichtung entstand.

In Experiment 2 zeigte sich weiterhin, dass nahe Ratschläge dazu führen, dass Menschen sich im Laufe eines Versuchsdurchgangs hinsichtlich ihrer Schätzungen sicherer fühlen. Wie bereits erwähnt, wird dieser Sicherheitszuwachs bei nahen Ratschlägen als Indikator für eine stattfindende soziale Validierung angesehen. Der leichte Anstieg der mittleren Sicherheiten bei den weitesten Ratschlagsentfernungen und die damit verbundene Signifikanz des quadratischen Kontrast trat wie bereits in Experiment 1 ein wenig unerwartet auf.

Hinsichtlich der Zusammenhänge zwischen der Schätzungsakkuratheit der Judges und den beiden AVn Ratschlagsgewichtung und Urteilssicherheit zeigten sich ähnli-

che Ergebnisse wie in Experiment 1. So haben erneut akkurater schätzende Probanden die ihnen dargebotenen Ratschläge insgesamt weniger genutzt, als Probanden mit weniger akkuraten Initialschätzungen. Weiterhin deutete sich in Experiment 2 an, dass die Gewichtungsverläufe akkurater schätzender Personen insgesamt weniger abhängig von der Ratschlagsdistanz sind, als die weniger akkurat schätzender Probanden.

6. Experiment 3

6.1 Zielsetzung und Hypothesen

Experiment 3 verfolgt drei Ziele: Erstens wird überprüft, ob die Auswirkungen der Ratschlagsdistanz stabil bleiben, wenn extremere Ratschlagsdistanzen als in Experiment 1 und 2 verwendet werden, denn es kann nicht ausgeschlossen werden, dass das bisher gefundene Muster der Beziehung zwischen Ratschlagsdistanz und -gewichtung bzw. Urteilssicherheit das Ergebnis des verwendeten Abweichungsrange ist. Aus den beiden Grafiken zum Verhältnis von Ratschlagsabweichung und Ratschlagsgewichtung der Experimente 1 und 2 (Abbildung 2 und 5) ist zu erkennen, dass die mittleren Ratschlagsgewichtungen bei den höchsten Ratschlagsdistanzen nicht wieder auf ein vergleichbar niedriges Niveau wie bei den sehr nahen Abweichungen zurückgehen. Dieser Umstand führt dazu, dass letztendlich auch immer der lineare Kontrast bei der Beschreibung des Zusammenhangs von Ratschlagsdistanz und -gewichtung Signifikanz erlangt. Auf Grund der sehr niedrigen Gewichtung bei nahen Ratschlägen und ihres darauffolgenden Anstiegs bei zunehmenden Ratschlagsdistanzen kennzeichnet dieser lineare Kontrast zudem einen ansteigenden Trend, das heißt, dass zu Beginn mit wachsender Ratschlagsentfernung auch dessen Gewichtung zunimmt, zumindest bis zu einem gewissen Punkt, ab dem die Gewichtung dann mit wachsender Ratschlagsdistanz wieder abnimmt. Dieser ansteigende Trend widerspricht nun den von Yaniv (2004b) dargelegten Ergebnissen in noch größerem Umfang, als die Signifikanz des von mir postulierten und gezeigten umgekehrt U-förmigen Kontrasts, der die Abnahme der Gewichtung mit wachsender Distanz zumindest teilweise enthält. Mit der Verwendung eines größeren Range von Ratschlagsabweichungen wurde in Experiment 3 überprüft, ob die Gewichtung von Ratschlägen bei sehr hohen Abweichungen irgendwann wieder auf ein vergleichbar (niedriges) Niveau wie bei sehr

nahen Abweichungen zurückgeht. Zweitens galt es zu untersuchen, ob die Distanzeffekte auf Ratschlagsgewichtung und Urteilssicherheit durch die Ratgeberkompetenz moderiert werden – beide Möglichkeiten, Moderation und Nicht-Moderation, sind hier jeweils denkbar. Aus den beiden vorangegangenen Experimenten wissen wir, welche Zusammenhänge zwischen Ratschlagsdistanz und Ratschlagsgewichtung bzw. Urteilssicherheit bei einem Ratgeber mit einem hohen Kompetenzniveau zu erwarten sind. Bei einem weniger kompetenten Ratgeber könnte nun einerseits der Distanzeffekt gänzlich obsolet werden, indem das Gewichtungs- und/oder Sicherheitsniveau grundsätzlich hier so niedrig ausfällt, dass kaum noch systematische Variationen mit der Ratschlagsdistanz möglich sind. Es ist aber auch denkbar, dass ein Distanzeffekt auch beim weniger kompetenten Ratgeber auftritt, aber der grafische Zusammenhang eine andere Form aufweist als beim kompetenten. So wäre in diesem Fall anzunehmen, dass die mittlere Ratschlagsgewichtung bei weit entfernten Ratschlagsdistanzen in deutlicherem Maße wieder abnimmt, als dies bisher beim kompetenten Ratgeber zu beobachten war. Bezüglich der Urteilssicherheit könnte die Signifikanz des quadratischen Kontrasts möglicherweise aufgehoben werden und ausschließlich ein linear fallender Trend zu beobachten sein. Andererseits besteht jedoch auch die Möglichkeit, dass die Ratgeberkompetenz sich weder auf die Existenz noch die Form des Distanzeffektes auswirkt, sondern lediglich auf das mittlere Gewichtungsniveau der Ratschläge und die durchschnittliche Urteilssicherheit der Probanden. Da bereits bekannt ist, dass die Expertise der Ratgeber das grundlegende Niveau der Ratschlagsgewichtung beeinflusst (z.B. Harvey & Fischer, 1997), und auch Bochner und Insko (1966) wissensspezifische Effekte im Hinblick auf den Distanzeffekt bei Einstellungsänderungen gefunden haben, erscheinen sie auch für meine Experimente nicht unwahrscheinlich. Experiment 3 überprüft daher drittens diese Möglichkeit. Viertens erfüllt Experiment 3 zusätzlich das Ziel, dass die Ergebnisse aus Experiment 1 und 2 trotz der enthaltenen Veränderungen repliziert werden. Es resultieren folgende (zum Teil bereits bekannte) Hypothesen:

Hypothese 1: Die Ratschlagsgewichtung ist eine quadratische (umgekehrt U-förmige) Funktion der Distanz zwischen Initialschätzung und Ratschlag, d.h. Ratschläge werden weniger stark gewichtet werden, wenn sie entweder sehr nah an oder sehr weit von der Initialschätzung des Judges entfernt sind, im Vergleich zu Ratschlägen,

deren Entfernung zur Initialschätzung moderat ausfällt (Replikation der Befunde aus Experiment 1 und 2).

Hypothese 2: Je näher Ratschläge an der Initialschätzung liegen, desto stärker wird die Urteilssicherheit zwischen Initial- und Finalurteil ansteigen (Replikation der Befunde aus Experiment 1 und 2).

Ich nehme außerdem an, dass sich auch im Rahmen meiner Experimente der etablierte Kompetenzeffekt auf die Ratschlagsgewichtung (z.B. Harvey & Fischer, 1997) zeigt. Daraus resultiert folgende Hypothese:

Hypothese 3: Ratschläge eines vermeintlich kompetenten Ratgebers werden stärker gewichtet als Ratschläge eines vermeintlich weniger kompetenten Ratgebers.

Hinsichtlich der Urteilssicherheit existieren nicht so klare Befunde im Zusammenhang mit der ausgewiesenen Ratgeberkompetenz. Budescu, Rantilla, Yu und Karelitz (2003) haben allerdings gezeigt, dass die Finalsicherheit mit zunehmender Ratgeberakkuratheit ansteigt. Zwar wird im Rahmen meiner Experimente kein Online-Feedback über die Akkuratheit der Ratschläge oder Finalschätzungen erteilt, jedoch werden die Probanden vor der „Ziehung“ des vermeintlichen Ratgebers dahingehend informiert, dass sich sein Rang über die Akkuratheit seiner Schätzungen ergeben hat. Dementsprechend kann auch für die AV Urteilssicherheit eine Hypothese zum Kompetenzeffekt formuliert werden:

Hypothese 4: Ratschläge eines vermeintlich kompetenten Ratgebers führen im Mittel zu einem größeren Anstieg der Urteilssicherheit als Ratschläge eines vermeintlich weniger kompetenten Ratgebers.

Die Untersuchung möglicher Interaktionen zwischen der Ratgeberkompetenz und der Art und Weise des Zusammenhangs zwischen der Ratschlagsdistanz und -gewichtung bzw. der Veränderung der subjektiven Sicherheit erfolgt mittels konkurrierender Hypothesen, da –wie bereits oben ausgeführt– keinen diesbezüglichen Annahmen ein theoretisch gerechtfertigter Vorzug gegeben werden kann.

Hypothese 5a: Im Hinblick auf die Ratschlagsgewichtung existiert keine signifikante Interaktion zwischen der Ratgeberkompetenz und Ratschlagsdistanz. Das heißt, die manipulierte Ratgeberkompetenz wirkt sich nicht systematisch auf die Existenz eines Distanzeffektes oder die Form des grafischen Zusammenhangs zwischen Ratschlagsdistanz und -gewichtung aus.

Hypothese 5b: Im Hinblick auf die Ratschlagsgewichtung existiert eine signifikante Interaktion zwischen der Ratgeberkompetenz und Ratschlagsdistanz. Das heißt, die manipulierte Ratgeberkompetenz wirkt sich systematisch auf die Existenz eines Distanzeffektes oder die Form des grafischen Zusammenhangs zwischen Ratschlagsdistanz und -gewichtung aus.

Hypothese 6a: Im Hinblick auf die Urteilssicherheit existiert keine signifikante Interaktion zwischen der Ratgeberkompetenz und Ratschlagsdistanz. Das heißt, die manipulierte Ratgeberkompetenz wirkt sich nicht systematisch auf die Existenz eines Distanzeffektes oder die Form des grafischen Zusammenhangs zwischen Ratschlagsdistanz und Urteilssicherheit aus.

Hypothese 6b: Im Hinblick auf die Urteilssicherheit existiert eine signifikante Interaktion zwischen der Ratgeberkompetenz und Ratschlagsdistanz. Das heißt, die manipulierte Ratgeberkompetenz wirkt sich systematisch auf die Existenz eines Distanzeffektes oder die Form des grafischen Zusammenhangs zwischen Ratschlagsdistanz und Urteilssicherheit aus.

6.2 Methode

6.2.1 Stichprobe und Design

40 Studenten der Universität Göttingen (*M* = 21.95 Jahre, *SD* = 2.6) nahmen an der Untersuchung teil; keiner von ihnen musste von der Datenanalyse ausgeschlossen werden. Die Stichprobe bestand aus 22 (55%) weiblichen und 18 männlichen Teilnehmern. Die Vergütung erfolgte erneut entweder monetär (6€) oder durch Vergabe von Versuchspersonenstunden. Auch in Experiment 3 wurde den Versuchspersonen mitgeteilt, dass die fünf akkuratesten Versuchspersonen Kino-Gutscheine im Wert von 15€

erhalten würden. Dies wiederum geschah, um eine grundlegende Motivation zur Abgabe präziser Schätzungen bei den Versuchsteilnehmern zu etablieren.

Das Design von Experiment 3 enthielt gegenüber den beiden vorangegangenen Experimenten zwei bedeutende Änderungen: Erstens erhielt die eine Hälfte der Stichprobe (n = 20) erneut Ratschläge eines vermeintlich kompetenten Ratgebers (mit dem Rang 7 unter 100 angeblichen früheren Teilnehmern), während die andere Hälfte (n = 20) der Stichprobe diesmal Ratschläge eines vermeintlich weniger kompetenten Ratgebers (mit dem Rang 78 unter 100 angeblichen früheren Teilnehmern) erhielt. Zweitens wurde der Innersubjektfaktor Ratschlagsdistanz um Abweichungen bis zu 150% erweitert. In Experiment 3 enthielt demnach ein mixed Design mit dem zweistufigen Zwischensubjektfaktor *Ratgeberkompetenz* (kompetent vs. weniger kompetent) und dem 30-stufigen Innersubjektfaktor *Ratschlagsdistanz* (5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 100%, 105%, 110%, 115%, 120%, 125%, 130%, 135%, 140%, 145%, 150%).

6.2.2 Ablauf

Den Aufgabeninhalt stellten in Experiment 3 erneut die Entfernungsschätzungen zwischen europäischen Hauptstädten dar. Der Ablauf eines Versuchsdurchgangs ist dementsprechend identisch zu Experiment 1. Die Abweichungen, welche der künstlich erzeugte Ratschlag zur Initialschätzung annehmen sollte, wurden in Experiment 3 zum Teil deutlich erhöht. Negative Abweichungen lagen wieder im Bereich von -5% bis -75% und wurden in 5%-Schritten gestuft. Das gleiche Abweichungsspektrum wurde auch für die positiven Abweichungen angewendet und um Abweichungen bis 150% erweitert, ebenfalls mit Abstufungen in 5%-Schritten. Jede positive und negative Abweichung von ±5% bis ±75% wurde einmal und jede (positive) Abweichung im Bereich von +80% bis +150% zweimal angewandt. Es resultierten daher wiederum 60 Versuchsdurchgänge, in denen die Ratschlagsdistanzen in derselben randomisierten Reihenfolge für alle Probanden angewendet wurden.

6.3 Ergebnisse und Diskussion

6.3.1 Analyse möglicher Störvariablen

Wie bereits bei den vorigen Experimenten, wurden nach der Berechnung der AT-Werte pro Trial solche außerhalb des Range von -1.5 bis +1.5 von der Analyse ausgeschlossen. Weiterhin wurden die Versuchsdurchgänge ausgeschlossen, in denen statt des manipulierten vermeintlichen Ratschlags Zufallszahlen dargeboten wurden. In Experiment 3 mussten dadurch 187 von insgesamt 2400 Versuchsdurchgängen ausgeschlossen werden, so dass die resultierende Ausschlussrate von knapp 8% deutlich geringer ausfiel als in Experiment 2. Erneut wurde auch das Alter der Probanden mit den vorhandenen AVn korreliert. Auch in Experiment 3 ergaben sich dabei erneut keine systematischen Korrelationen zwischen Alter und mittlerer Ratschlagsgewichtung oder Urteilssicherheit, beide *r*s(40) < .10, beide *p*s > .60. Bei Aufnahme des Faktors Geschlecht in die Hauptanalyse ergab sich auch dieses Mal kein signifikanter Haupteffekt, weder im Hinblick auf die Ratschlagsgewichtung, $F(1, 26) = 0.96$, $p = .34$, $\eta^2 = .04$, noch in Bezug auf die Veränderung der Urteilssicherheit, $F(1, 29) = 0.14$, $p = .71$, $\eta^2 = .005$. Des Weiteren gab es keine systematischen Interaktionen des Geschlechts der Probanden mit einer oder mehreren UVn, alle *F*s < 3.5 , alle *p*s > .07, alle $\eta^2 < .12$.

6.3.2 Ratschlagsgewichtung

Wie auch schon in den beiden vorangegangenen Experimenten, begann die Analyse der Ratschlagsgewichtung mit der Berechnung der AT-Werte für jeden Trial bei jedem einzelnen Versuchsteilnehmer. Im Anschluss erfolgte die Umordnung dieser ermittelten AT-Scores gemäß der im jeweils betreffenden Versuchsdurchgang angewandten absoluten Ratschlagsabweichung sowie der Zwischensubjektbedingung Advisorkompetenz. Die Analyse erfolgte mit einer 2 (*Kompetenz des Ratgebers:* kompetent vs. weniger kompetent) x 30 (*Ratschlagsdistanz:* 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 100%, 105%, 110%, 115%, 120%, 125%, 130%, 135%, 140%, 145%, 150%)-ANOVA mit der Art des Ratgebers als Zwischensubjekt- und der Ratschlagsabweichung als Innersubjektfaktor. Hierbei ergab sich ein signifikanter Effekt für den Zwischensubjektfaktor Ratgeberkompetenz, $F(1, 28) = 8.66$, $p = .006$, $\eta^2 = .24$. In Übereinstimmung mit Hypothese 3 wurden

Ratschläge im Mittel stärker gewichtet, wenn sie von einem vermeintlich kompetenten Ratgeber erteilt wurden (*M* = .26, *SD* = .22) als von einem angeblich weniger kompetenten Ratgeber (*M* = .14, *SD* = .16). Hypothese 3 wird daher angenommen. Des Weiteren erwies sich auch in Experiment 3 der Innersubjektfaktor Ratschlagsdistanz als signifikant, $F(11.13, 311.64) = 3.75$, $p < .001$, $\eta^2 = .12$, wurde jedoch nicht von der Ratgeberkompetenz moderiert, $F(11.13, 311.64) = 0.55$, $p = .87$, $\eta^2 = .02$. Demnach scheint die Kompetenz eines Ratgebers keinen Einfluss auf die Art der Beziehung zwischen Ratschlagsdistanz und –gewichtung auszuüben, lediglich auf das Ausmaß des mittleren Gewichtungsniveaus, wie auch an Hand von Abbildung 12 deutlich zu erkennen ist. Hypothese 5a kann daher als eine der konkurrierenden Hypothesen angenommen werden. Um die Art der Beziehung zwischen Ratschlagsdistanz und –gewichtung genauer zu explorieren, wurden erneut die Kontraste berechnet. Hypothesengemäß erlangte dabei der quadratische Kontrast Signifikanz, $F(1, 28) = 19.44$, $p < .001$, $\eta^2 = .41$. Die Hypothese 1 wird somit angenommen. Zudem wurde erneut auch der lineare Kontrast statistisch signifikant, $F(1, 28) = 7.42$, $p = .011$, $\eta^2 = .21$.

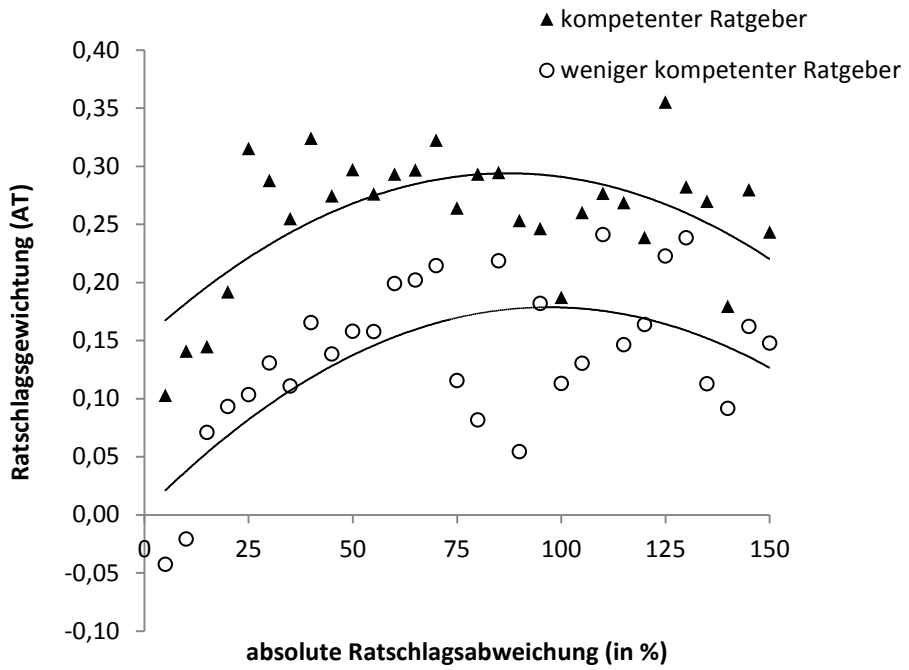

Abbildung 12

grafische Beziehung zwischen Ratschlagsdistanz (in %) und mittlerer Ratschlagsgewichtung für den kompetenten und weniger kompetenten Ratgeber in Experiment 3

Bei Betrachtung von Abbildung 12 fällt analog zu den vorangegangenen Experimenten auf, dass Ratschläge weniger stark gewichtet wurden, wenn sie relativ nah an der Initialschätzung lagen. In der Bedingung, in der Probanden die Ratschläge eines vermeintlich kompetenten Ratgebers erhielten, resultierte für die Abweichungen von 5% bis 15% ein mittlerer AT-Wert von .13, während bei den übrigen Abweichungen von 20% bis 150% ein Wert von .27 resultierte. In der Bedingung mit den Ratschlägen eines angeblich weniger kompetenten Ratgebers ergaben sich für die geringen Abweichungen mittlere Gewichtungen von im Mittel .002 und für die höheren Abweichungen ein Wert von .15.

Wie sich bereits in den ersten beiden Experimenten angedeutet hat, drücken die Mittelwerte vermutlich nicht das grundlegende Gewichtungsniveau bei den einzelnen Ratschlagsdistanzen aus, sondern kommen dadurch zu Stande, dass mit unterschiedlichen Abweichungen des Ratschlags von der Initialschätzungen in unterschiedlicher Häufigkeit vor allem Null-Gewichtungen, 50%-Gewichtungen und Gewichtungen um einen Grundbetrag von 20% bis knapp 40% angewendet werden. Im Anschluss erfolgte daher wieder die Analyse der Auftretenshäufigkeiten bestimmter Gewichtungskategorien insgesamt und in Abhängigkeit der manipulierten Ratschlagsdistanzen. Bezüglich der Gesamthäufigkeiten der Gewichtungskategorien zeigte sich für den kompetenten Ratgeber keine so deutliche zwei-gipflige Verteilung der Gewichtungen wie in den beiden vorangegangenen Experimenten. Geringfügige Gewichtungen über 0% bis knapp 40% wurden insgesamt deutlich häufiger vorgenommen, als in den vorangegangen Experimenten und die *Averaging*-Strategie etwas seltener.

Wie es der Haupteffekt der Ratgeberkompetenz bereits anzeigte, wurden Ratschläge eines vermeintlich weniger kompetenten Ratgebers im Mittel insgesamt weniger gewichtet als die eines kompetenten Ratgebers. Die Aufschlüsselung der Gewichtungsstrategien ergab in diesem Zusammenhang, dass die Ratschläge des weniger kompetenten Ratgebers in der Mehrzahl der Fälle vollständig diskontiert wurden, was zu der niedrigeren Gesamtgewichtung dieser Ratschläge gegenüber den Ratschlägen des kompetenteren Ratgebers geführt haben dürfte. Die daneben vorkommenden Gewichtungsausmaße bewegten sich mehrheitlich in einem Bereich bis knapp 40%. Die

Bildung eines Mittelwertes aus der eigenen Schätzung und dem Ratschlag wurde deutlich seltener vorgenommen als beim kompetenten Ratgeber. Die Verteilung der Gewichtungen zeigt daher einen eher rechtsschiefen Verlauf (siehe Abbildung 13).

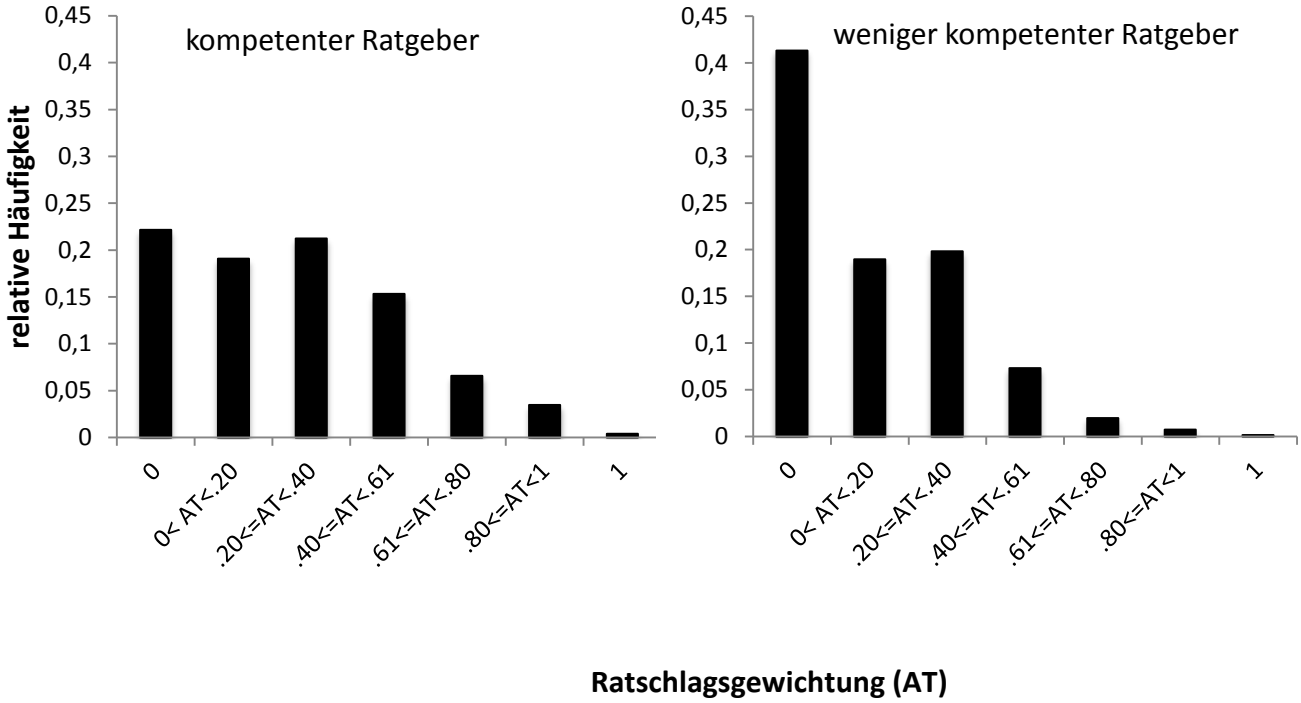

Abbildung 13

Verteilung der Ratschlagsgewichtungen in Experiment 3, getrennt dargestellt für den kompetenten und den weniger kompetenten Ratgeber

Um genaueren Aufschluss über das Zustandekommen der umgekehrten U-Form des Distanzeffekts der Gewichtung zu erlangen, wurden im nächsten Schritt wieder die Gewichtungshäufigkeiten in Abhängigkeit der Ratschlagsdistanzen analysiert. In Abbildung 14 sind diese Verläufe dargestellt. Da keine signifikante Interaktion zwischen manipulierter Ratgeberkompetenz und der Ratschlagsdistanz vorlag, wurde auf eine nach Kompetenzstufen getrennte Darstellung verzichtet.

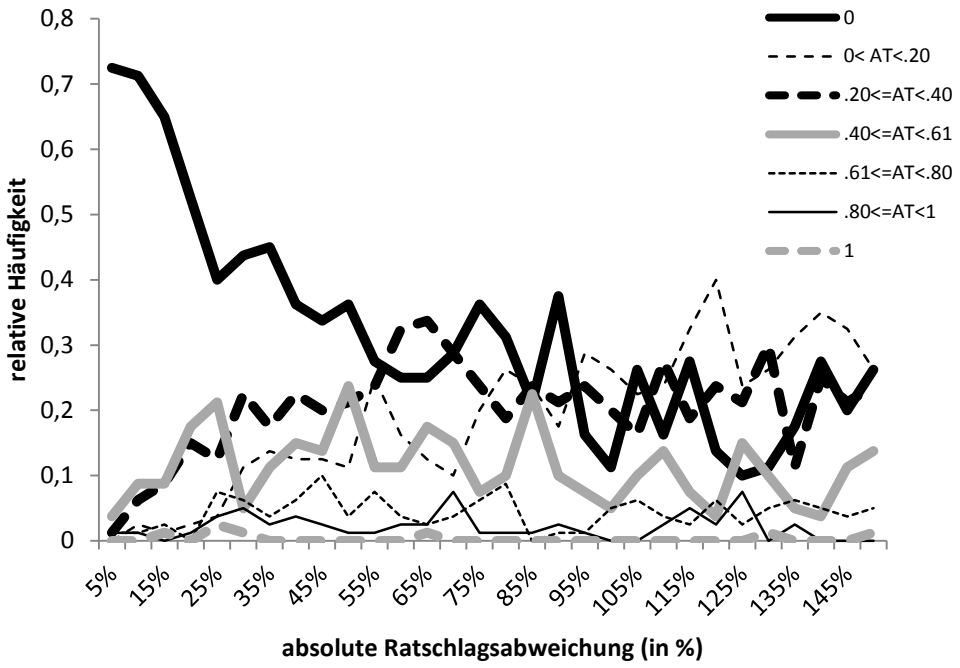

Abbildung 14

relative Häufigkeiten der Ratschlagsgewichtungen in Abhängigkeit der absoluten Ratschlagsentfernung in Experiment 3

Wie auch in Experiment 1 und 2 wurden die sehr nahen Ratschläge zumeist vollständig diskontiert oder aber, wenngleich deutlich seltener, der Mittelwert aus der eigenen Schätzung und dem erhaltenen Ratschlag gebildet. Der hohe Anteil an Nullgewichtungen bei sehr nahen Ratschlägen zeichnet sich damit wiederum fast allein verantwortlich für die geringen mittleren Gewichtungskoeffizienten bei diesen Distanzen. Mit zunehmenden Ratschlagsdistanzen nimmt die Häufigkeit der Nullgewichtungen jedoch fast kontinuierlich ab. Moderat entfernten Ratschläge wurde sich so vorwiegend in einem Ausmaß bis etwa 40% angenähert, was die höheren AT-Mittelwerte gegenüber den nahen Ratschlagsabweichungen und den Anstieg der Gewichtungskurve bedingt. Mit stärker zunehmenden Distanzen wurden Ratschläge dann überwiegend nur noch in einem geringfügigen Umfang bis knapp unter 20% gewichtet – die AT-Mittelwerte fielen dadurch wieder geringer aus mit der Folge des Abfalls der Gewichtungskurve, wenngleich auch nicht so gering wie bei nahen Distanzen, da der Anteil der Nullgewichtungen deutlich unter dem Anfangsniveau lag. Diese Kategorie der Gewichtungen bis 20% spielte in den vorangegangenen Experimenten meist keine er-

wähnenswerte Rolle, was vermutlich daran liegt, dass sie mit großer Häufigkeit erst ab Ratschlagsentfernungen von 80% und mehr auftrat, welche über die in den Experimenten 1 und 2 manipulierten Distanzen hinausging.

6.3.3 Urteilssicherheit

Wie bei den vorangegangenen Analysen zur Urteilssicherheit wurde damit begonnen, die Differenzen der Urteilssicherheiten pro Proband und Versuchsdurchgang zu berechnen. Im Anschluss wurden diese wieder pro Versuchsteilnehmer und absoluter prozentualer Abweichung gemittelt und entsprechend dieser Abweichungen angeordnet. Die zweifaktorielle rm-ANOVA mit 30-stufiger Messwiederholung auf dem Faktor Ratschlagsdistanz und dem Zwischensubjektfaktor Ratgeberkompetenz zeigte auch bei der AV der Urteilssicherheit einen signifikanten Haupteffekt für die Ratgeberkompetenz, $F(1, 31) = 4.97$, $p = .033$, $\eta^2 = .14$: Versuchsteilnehmer erlangten einen durchschnittlich höheren Sicherheitszuwachs im Laufe eines Versuchsdurchgangs, wenn der Ratschlag von einem vermeintlich kompetenten Ratgeber stammte ($M = .32$, $SD = .76$), als wenn sie ihn von einem angeblich weniger kompetenten Ratgeber erhielten ($M = .09$, $SD = .56$). Damit kann Hypothese 4 angenommen werden. Im Falle des weniger kompetenten Ratgebers zeigten sich über einen weiten Bereich von Distanzen zudem negative mittlere Sicherheitsveränderungen (siehe Abbildung 15). Das heißt, dass die Probanden durch die Ratschläge eines weniger kompetenten Ratgebers bei einigen Abweichungen eher verunsichert wurden.

Außerdem ergab sich wiederum ein signifikanter Haupteffekt für den Innersubjektfaktor Ratschlagsdistanz, $F(11.79, 365.54) = 4.62$, $p < .001$, $\eta^2 = .13$. Eine signifikante Interaktion zwischen den beiden Faktoren Kompetenz und Distanz trat nicht auf, $F(11.79, 365.54) = 1.196$, $p = .28$, $\eta^2 = .03$, das heißt die Art der (grafischen) Beziehung zwischen Ratschlagsdistanz und Veränderungen der subjektiven Urteilssicherheit ist unabhängig von der Kompetenz eines Ratgebers und dadurch auch in gewisser Weise der Qualität der Ratschläge. Hypothese 6a kann dadurch angenommen werden. Bezüglich des Distanzeffektes, und zur Prüfung von Hypothese 2, wurden erneut sowohl der lineare als auch der quadratische Kontrast auf Signifikanz überprüft. Es zeigte sich sowohl eine Signifikanz des linearen, $F(1, 31) = 32.90$, $p < .001$, $\eta^2 = .52$, als auch des

quadratischen Kontrasts, $F(1, 31) = 15.95$, $p < .001$, $\eta^2 = .34$. Dies deutet auf einen sinkenden Sicherheitszuwachs bei ansteigender Ratschlagsdistanz hin, bis hin zu einem gewissen Wendepunkt, ab dem Menschen sich wieder sicherer bezüglich ihrer Schätzungen werden (siehe Abbildung 15). Dennoch zeigte sich, dass im Mittel die Urteilssicherheiten mit wachsender Ratschlagsdistanz deutlich abnahmen. In der Bedingung, in der die Probanden Ratschläge von angeblich weniger kompetenten Versuchspersonen erhielten, zeigte sich für die ersten drei Abweichungstufen (5%, 10% und 15%) ein mittlerer Zuwachs an Urteilssicherheit von .66. Bei Abweichungen von 20% und mehr fiel der mittlere Sicherheitszuwachs merklich geringer aus und erreichte bei einigen Ratschlagsabweichungen im Mittel negative Werte. Der Mittelwert lag in diesem Bereich (20% Abweichung bis 150% Abweichung) bei .03. Erhielten die Versuchsteilnehmer Ratschläge eines vermeintlich kompetenten Ratgebers, betrug der mittlere Zuwachs an Urteilssicherheit bei den nahen Abweichungen im Mittel .80. Bei Abweichungen von 20% bis 150% ergab sich dann ein Mittelwert von .27.

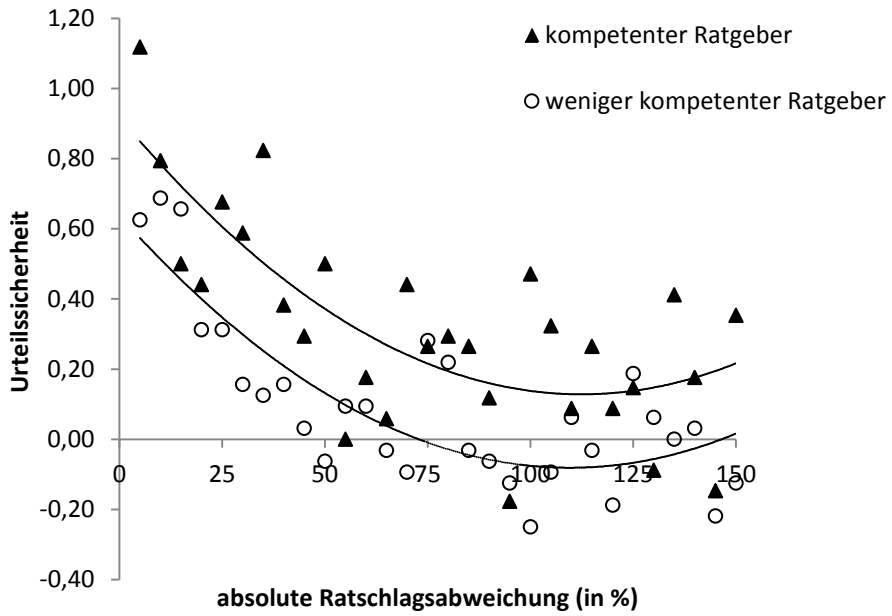

Abbildung 15

grafische Beziehung zwischen Ratschlagsdistanz (in %) und Urteilssicherheit für den kompetenten und weniger kompetenten Ratgeber in Experiment 3

Erneut wurde die Abhängigkeit der beiden untersuchten AVn von der Ratschlagsdistanz jeweils durch gegenläufige Kontraste beschrieben – ob nun linear stei-

gend für die Ratschlagsgewichtung und fallend für die Urteilssicherheit oder quadratisch mit einer umgekehrten U-Form im Falle der Ratschlagsgewichtung und einer U-Form für die Urteilssicherheit. Dementsprechend fallen die Korrelationen der jeweiligen Mittelwerte der AVn pro absoluter Ratschlagsabweichung sowohl für den kompetenten als auch für den weniger kompetenten Ratgeber signifikant negativ aus, $r_{komp}(30) = -.44$, $p = .015$ und $r_{\neg komp}(30) = -.61$, $p < .001$. Der Verlauf der mittleren Ratschlagsgewichtungen und Urteilssicherheiten in Abhängigkeit der ihnen vorausgehenden Ratschlagsdistanzen kann bei Kenntnis des Verlaufs einer der beiden Variablen gut vorhergesagt werden. Auf Ebene der einzelnen Probanden ist dieser Zusammenhang jedoch nicht so eindeutig. In der Bedingung mit einem vermeintlich kompetenten Ratgeber beträgt die mittlere Korrelation hier $r = .04$ und bei einem angeblich weniger kompetenten $r = -.05$. Bei Betrachtung der individuellen Korrelationen von Gewichtung und Sicherheit ergibt sich keine derartige Systematik, so kommen in keiner der beiden Experimentalbedingungen negative oder positive Korrelationen signifikant häufiger vor, $\chi^2(1, N = 39) = 2.06$, $p = .15$. Dementsprechend kann auch hier wieder der Umstand vorgelegen haben, dass die bei steigender Urteilsadjustierung in Richtung des Ratschlags abnehmende Urteilssicherheit betragsmäßig stärker ausfällt als eventuelle Zunahmen und daher den Mittelwert in die entsprechende Richtung beeinflusst.

6.3.4 Explorative Datenanalyse: Schätzungsakkuratheit der Judges

Auf die Frage, welchen Platz von 100 Teilnehmern in diesem Experiment sie ihrer eigenen Leistung zuteilen würden, gaben die Probanden ihren eigenen Schätzungen im Mittel einen Rang von 50 ($M = 50.39$, $SD = 28.83$) – zeigten also keinen diesbezüglichen Überoptimismus. Wird die subjektive Kompetenzbeurteilung der Judges in Abhängigkeit ihrer Bedingungszugehörigkeit analysiert, zeigt sich etwas Interessantes; und zwar schätzen die Probanden, die zuvor die Ratschläge eines vermeintlich kompetenten Ratgebers erhalten, ihre eigene Kompetenz signifikant höher bzw. besser ein als Probanden, die Ratschläge des angeblich weniger kompetenten Ratgebers erhielten, $t(36) = 2.72$, $p = .01$, $d = .89$. Konkret bedeutet dies, dass Probanden, die Ratschläge eines Ratgebers mit dem Rang 7 erhielten, sich und ihren Leistungen durchschnittlich den Rang 38 ($M = 38$, $SD = 30.72$) zuwiesen, während die Versuchsteilnehmer in der Bedingung mit einem Ratgeber mit einem 78. Rang, sich durchschnittlich auf Platz 61

sahen (*M* = 61.55, *SD* = 22.32). Dies deutet darauf hin, dass die Probanden in Experiment 3 hinsichtlich der Einschätzung ihrer eigenen Leistungen möglicherweise einem Anker-Effekt (Tversky & Kahnemann, 1974) erlegen sind.

In der Bedingung mit einem angeblich kompetenten Advisor zeigte sich weiterhin eine signifikante Korrelation zwischen der vermuteten eigenen Kompetenz der Judges und der mittleren Ratschlagsgewichtung, $r(18) = .50$, $p = .03$. Das heißt, dass mit abnehmender subjektiver Kompetenz, sprich, steigenden Rangplätzen, eine höhere Gewichtung der Ratschläge einhergeht. Systematische Zusammenhänge zwischen subjektiver Urteilsgüte und der mittleren Veränderung der Urteilssicherheit traten nicht auf, $r(18) = -.22$, $p = .37$. In der Bedingung mit einem vorgeblich weniger kompetenten Ratgeber gab es keine signifikanten Zusammenhänge zwischen dem vermuteten Rangplatz und der mittleren Ratschlagsgewichtung, $r(20) = .23$, $p = .33$, oder der Urteilssicherheit, $r(20) = -.44$, $p = .05$, wobei die letztere Korrelation betragsmäßig doch recht hoch ausfiel und die Schwelle der Signifikanz nur knapp verfehlte. Eine signifikante Korrelation zwischen subjektiver (Rangplätze) und objektiver Leistungsbeurteilung (MAPE) bestand im Übrigen nicht, $r(38) = .26$, $p = .12$.

Die Analyse objektiver Maße für die Schätzungsakkuratheit der Judges zeigte eine systematische Überschätzung der Entfernungen zwischen zwei europäischen Städten um ca. 37% an ($M_{MPE} = .37$, $SD = .45$), $t(39) = 5.19$, $p < .001$, $d = .82$. Wurden die Probanden an Hand des MAPE-Medians anschließend in zwei Gruppen zu akkuraten und weniger akkuraten Versuchspersonen eingeteilt und dieser Faktor als zusätzlicher Zwischensubjektfaktor neben der *Ratgeberkompetenz* in die Hauptanalyse integriert, zeigte sich in diesem Experiment kein Haupteffekt der Schätzungsakkuratheit, $F(1, 26) = 0.33$, $p = .57$, $\eta^2 = .01$. Die mittlere Gewichtung von Ratschlägen durch akkurate Versuchspersonen ($M = .19$, $SD = .19$) unterschied sich in Experiment 3 also nicht signifikant von der bei weniger akkuraten Versuchspersonen ($M = .21$, $SD = .22$), bzw. anders als in den beiden vorangegangenen Experimenten gewichteten weniger akkurate Versuchspersonen die ihnen dargebotenen Ratschläge nicht systematisch stärker als bessere Versuchspersonen. Eine signifikante Interaktion zwischen der Ratgeberkompetenz und der Schätzungsakkuratheit lag dabei nicht vor, $F(1, 26) = 0.89$, $p = .67$, $\eta^2 <$

.01, so dass nicht davon auszugehen ist, dass die beiden Ratgeberbedingungen sich hinsichtlich des Zusammenhangs zwischen der Schätzungsakkuratheit der Judges und der mittleren Ratschlagsgewichtung voneinander unterscheiden. Die signifikante Abhängigkeit der Ratschlagsgewichtung von der Ratschlagsentfernung, $F(10.63, 276.26) = 3.58$, $p < .001$, $\eta^2 = .12$, unterscheidet sich nicht für die beiden Akkuratheitsgruppen, das heißt, die Interaktion zwischen Ratschlagsdistanz und Schätzungsakkuratheit wurde nicht signifikant, $F(10.63, 276.26) = 1.29$, $p = .24$, $\eta^2 = .05$. Es ist somit davon auszugehen, dass für unterschiedlich akkurate Versuchspersonen dieselbe Beziehung zwischen Ratschlagsdistanz und –gewichtung angenommen werden kann. Erneut zeigt sich auch die relativ niedrige Gewichtung naher Ratschläge (bis 15% Abweichung) gegenüber den übrigen weiter entfernten für beide Akkuratheitsgruppen. So gewichten akkurate Versuchspersonen Ratschläge bis 15% Abweichung im Mittel zu .04 und die übrigen zu etwa .20 und weniger akkurate Versuchspersonen nahe Ratschläge zu .10 und weiter entfernte zu durchschnittlich .23.

Die Urteilssicherheit wurde abschließend auch auf mögliche Abhängigkeiten von der Schätzungsakkuratheit der Judges überprüft. Dabei zeigte sich weder ein signifikanter Haupteffekt für die Akkuratheit, $F(1, 29) = 0.11$, $p = .74$, $\eta^2 < .01$, noch eine Interaktion der Akkuratheit mit der UV Ratgeberkompetenz, $F(1, 29) = 0.41$, $p = .53$, $\eta^2 = .01$, oder dem Innersubjektfaktor Ratschlagsdistanz, $F(11.16, 323.63) = 0.88$, $p = .56$, $\eta^2 = .03$. Akkurate Versuchspersonen und weniger akkurater Versuchspersonen unterschieden sich also im Mittel nicht hinsichtlich des Zuwachses an Urteilssicherheit im Laufe eines Trials ($M_{akkurat} = .19$, $SD = .67$ vs. $M_{\neg akkurat} = .23$, $SD = .71$), und dies unabhängig von der Kompetenzmanipulation bezüglich des Ratgebers. Auch der Distanzeffekt bei der Urteilssicherheit, $F(11.16, 323.63) = 4.48$, $p < .001$, $\eta^2 = .13$, zeigte demnach keine Abhängigkeit von der Schätzungsakkuratheit der Judges.

Zusammenfassend replizieren die Ergebnisse von Experiment 3 diejenigen von Experiment 1 und 2. Es konnte erneut die hypostasierte kurvilineare, umgekehrt U-förmige Beziehung zwischen Ratschlagsdistanz und –gewichtung gefunden werden. Außerdem konnte auch der kurvilineare, U-förmige Zusammenhang zwischen Ratschlagsdistanz und der Veränderung der subjektiven Urteilssicherheit neuerlich gefun-

den werden. Der Einfluss der Ratschlagsdistanz auf die Ratschlagsgewichtung sowie die Urteilssicherheit erwies sich in Experiment 3 als unabhängig von der vermeintlichen Ratgeberkompetenz. Zwar gibt es grundlegende Niveauunterschiede im Ausmaß der AVn bei unterschiedlich kompetenten Ratgebern, die sich in einer erwartungsgemäß höheren Gewichtung sowie in einem höheren Sicherheitszuwachs bei Ratschlägen des kompetenten gegenüber denen des weniger kompetenten widerspiegeln. Eine Interaktion zwischen Distanz und Kompetenz gab es jedoch nicht. Experiment 3 gibt dadurch erste Hinweise darauf, dass die Ratgeberkompetenz den Distanzeffekt nicht moderiert. Die Verwendung größerer Ratschlagsabweichungen führte ferner nicht zum Rückgang der AT-Mittelwerte auf das Anfangsniveau, da auch bei sehr hohen Abweichungen die mittlere Gewichtung von Ratschläge gegenüber sehr nahen Ratschlägen noch vergleichsweise hoch ausfiel.

Die Analyse der Gewichtungsverteilung für die einzelnen Distanzen offenbarte in diesem Zusammenhang, dass bei sehr nahen Ratschlägen ein vergleichsweise hoher Anteil von Nullgewichtungen der Ratschläge an allen Gewichtungskategorien vorlag, welcher bei den weit entfernten Ratschlägen deutlich kleiner ausgeprägt war. Stattdessen werden hier mehrheitlich Gewichtungen bis .20, aber auch teilweise bis .40 vorgenommen, sodass im Mittel bei diesen Distanzen wiederum eine höhere Durchschnittsgewichtung ermittelt werden konnte als bei den sehr geringen Abweichungen. Gegenüber den vorangegangen Experimenten zeigte sich eine Verschiebung der Gesamthäufigkeiten in Richtung der geringfügigen Gewichtungen bis 20% oder bis 40%, welche in Experiment 3 häufiger vorgenommen wurden als die Bildung eines Mittelwertes aus Initialschätzung und dem erhaltenen Ratschlag.

Die Auswirkungen der judge-eigenen Kompetenz auf die Ratschlagsgewichtung waren etwas anderer Gestalt als in den beiden vorangegangenen Experimenten. So gab es im Mittel keine Gewichtungsunterschiede zwischen den beiden Gruppen unterschiedlich akkurat schätzender Personen. Auch die Urteilssicherheit zeigte sich erneut unbeeinflusst von der aufgabenbezogenen Kompetenz der Versuchspersonen.

7. Experiment 4

7.1 Zielsetzung und Hypothesen

Bis hierhin kann festgehalten werden, dass in drei Experimenten durchgängig die zu Beginn postulierte kurvilineare, umgekehrt U-förmige Beziehung zwischen Ratschlagsdistanz und –gewichtung gefunden werden konnte. Diese Beziehung trat zudem unabhängig vom verwendeten Aufgabentyp und der Ratgeberkompetenz auf. Auf Grund dieser stabilen Befundlage wird davon ausgegangen, dass die quadratische Beziehung zwischen Distanz und Ratschlagsgewichtung einen relativ weiten Geltungsbereich besitzt. Die hier berichteten Ergebnisse widersprechen damit den von Yaniv (2004b) getroffenen Annahmen und seinen Befunden: Während er einen monoton fallenden Trend für die Beziehung zwischen Ratschlagsdistanz und –gewichtung annimmt und nachweisen konnte, konnte in den hier berichteten Experimenten konstant die postulierte quadratische, umgekehrt U-förmige Beziehung gefunden werden. Es gibt jedoch einen wesentlichen Unterschied methodischer Natur zwischen seinem und dem hier berichteten experimentellen Vorgehen: In den Experimenten 1 bis 3 wurde die Ratschlagsdistanz kontinuierlich in 5%-Schritten manipuliert bzw. alterniert. Diese Vorgehensweise ermöglicht eine relativ detaillierte Erforschung der Beziehung zwischen der Ratschlagsdistanz und den beiden abhängigen Variablen, da der Wertebereich der Ratschlagsdistanz relativ engmaschig abgedeckt wird. Yaniv (2004b) hingegen bot seinen Probanden Ratschläge dar, die hinsichtlich ihrer Abweichungen von den Initialschätzungen der Versuchsteilnehmer in drei große Gruppen kategorisiert werden können. Wie bereits an anderer Stelle erwähnt, handelte es sich bei diesen Kategorien um Ratschläge, die entweder „nah“, „intermediär“ oder „weit“ von der Initialschätzung entfernt lagen. Obwohl die Ratschläge innerhalb dieser Kategorien in einem gewissen Ausmaß variierten, könnte diese kategoriale Manipulation dazu geführt haben, dass Yanivs Versuchsteilnehmer deutlichere Unterschiede zwischen nahen, moderat entfernten und weit entfernten Ratschlägen wahrnahmen und dementsprechend diese auch anders gewichtet haben. Die Überprüfung dieser Möglichkeit erfolgte im Rahmen von Experiment 4, in dem ein Vergleich zwischen kontinuierlich und kategorial manipulierten Ratschlägen stattfand. Zusätzlich kann dadurch untersucht werden, ob die kon-

tinuierliche anstelle einer kategorialen Ratschlagsmanipulation zu dem nur gering ausgeprägten Abfall der AT-Werte bei weit entfernten Ratschlägen bei den bereits berichteten Experimenten geführt hat.

Experiment 4 wurde daher durchgeführt, um zu überprüfen, ob die konsistent über drei Experimente hinweg gezeigten Ergebnisse (teilweise) durch die Art und Weise der Manipulation des Abstandes zwischen Initialurteil und Ratschlag zu Stande gekommen sind. Da es aber keine theoretische Grundlage gibt, auf Grund derer Annahmen über mögliche Effekte der Art der Distanzmanipulation auf die Ratschlagsgewichtung und Urteilssicherheit sowie insbesondere mögliche Interaktionen mit den anderen UVn getroffen werden können, erfolgt die Analyse zu möglichen Interaktionseffekten dieser Variable mit der UV Ratschlagsdistanz auf Grundlage konkurrierender Hypothesen. Es können jedoch folgende bereits bekannte Hypothesen postuliert werden.

Hypothese 1: Die Ratschlagsgewichtung ist eine quadratische (umgekehrt U-förmige) Funktion der Distanz zwischen Initialschätzung und Ratschlag, d.h. Ratschläge werden weniger stark gewichtet werden, wenn sie entweder sehr nah an oder sehr weit von der Initialschätzung des Judges entfernt sind, im Vergleich zu Ratschlägen, deren Entfernung zur Initialschätzung moderat ausfällt (Replikation der Befunde aus Experiment 1 bis 3).

Hypothese 2: Je näher Ratschläge an der Initialschätzung liegen, desto stärker wird die Urteilssicherheit zwischen Initial- und Finalurteil ansteigen (Replikation der Befunde aus Experiment 1 bis 3).

Hypothese 3: Ratschläge eines vermeintlich kompetenten Ratgebers werden stärker gewichtet als Ratschläge eines vermeintlich weniger kompetenten Ratgebers (Replikation der Befunde aus Experiment 3).

Hypothese 4: Ratschläge eines vermeintlich kompetenten Ratgebers führen zu einem durchschnittlich größeren Anstieg der Urteilssicherheit als Ratschläge eines vermeintlich weniger kompetenten Ratgebers (Replikation der Befunde aus Experiment 3).

Hypothese 5: Der gefundene Zusammenhang zwischen der Distanz von Initialschätzung und Ratschlag und der Ratschlagsgewichtung ist unabhängig von der ausgewiesenen Ratgeberkompetenz (Replikation des Befundes aus Experiment 3).

Hypothese 6: Der gefundene Zusammenhang zwischen der Distanz von Initialschätzung und Ratschlag und der Urteilssicherheit ist unabhängig von der ausgewiesenen Ratgeberkompetenz (Replikation des Befundes aus Experiment 3).

Hinsichtlich der Art der Distanzmanipulation und deren Auswirkungen auf den Distanzeffekt werden folgende konkurrierende Hypothesen aufgestellt:

Hypothese 7a: Die Art der Distanzmanipulation wirkt sich nicht systematisch auf den Zusammenhang zwischen Ratschlagsdistanz und –gewichtung aus. Das heißt in beiden Distanzmanipulationsbedingungen besteht der gleiche (umgekehrt U-förmige) Zusammenhang zwischen Ratschlagsdistanz und Ratschlagsgewichtung.

Hypothese 7b: Die Art der Distanzmanipulation wirkt sich systematisch auf den Zusammenhang zwischen Ratschlagsdistanz und –gewichtung aus. Das heißt in beiden Distanzmanipulationsbedingungen besteht ein unterschiedlicher grafischer Zusammenhang zwischen Ratschlagsdistanz und Ratschlagsgewichtung.

Die gleichen konkurrierenden Hypothesen werden auch im Hinblick auf einen Interaktionseffekt zwischen der Distanzmanipulation und der Ratschlagsdistanz für AV der Urteilssicherheit formuliert:

Hypothese 8a: Die Art der Distanzmanipulation wirkt sich nicht systematisch auf den Zusammenhang zwischen Ratschlagsdistanz und Urteilssicherheit aus. Das heißt in beiden Distanzmanipulationsbedingungen besteht der gleiche (lineare und/oder U-förmige) Zusammenhang zwischen Ratschlagsdistanz und Urteilssicherheit.

Hypothese 8b: Die Art der Distanzmanipulation wirkt sich systematisch auf den Zusammenhang zwischen Ratschlagsdistanz und Urteilssicherheit aus. Das heißt in bei-

den Distanzmanipulationsbedingungen besteht ein unterschiedlicher grafischer Zusammenhang zwischen Ratschlagsdistanz und Urteilssicherheit.

7.2 Methode

7.2.1 Stichprobe und Design

Insgesamt wurden für Experiment 4 die Daten von 105 Personen erhoben. Neun Personen mussten jedoch entweder auf Grund von Teilnahme an vorhergehenden vergleichbaren Untersuchungen bei gleichzeitiger Nicht-Beantwortung des Suspicion Checks (n = 3), oder Antworten im Suspicion Check, die darauf schließen lassen, dass die Probanden die Manipulation durchschaut hatten oder zumindest an der Echtheit der Ratschläge zweifeln (n = 4), oder aber auf Grund von nicht sinnvoll auswertbaren Daten nach Programminstabilität (n = 1) und auffälligem Antwortverhalten[7] (n = 1) ausgeschlossen werden. Die Daten der verbleibenden 96 Studierenden (M_{Alter} = 21.36 Jahre, *SD* = 2.65) wurden in die Analyse integriert. 65 (67%) der Versuchsteilnehmer waren weiblich und 32 männlich. Sie erhielten die gleiche Vergütung in Form von Versuchspersonenstunden oder Geld (6€) wie bereits die Probanden der vorangegangenen Experimente. Zusätzliche Motivationsanreize wurden erneut mit der Ankündigung, Kino-Gutscheine im Wert von 15 Euro unter den Top-fünf der Versuchsteilnehmer zu verlosen, bereit gestellt.

Das Design von Experiment 4 wurde im Hinblick auf die mit ihm verfolgten Untersuchungsziele wie folgt angepasst: Neben dem bereits in Experiment 3 angewendeten Zwischensubjekfaktor *Ratgeberkompetenz* (kompetent vs. weniger kompetent) wurde jetzt ein weiterer Zwischensubjektfaktor ergänzt. Dieser Faktor wurde mit *Distanzmanipulation* (kategorial vs. kontinuierlich) betitelt und legte fest, auf welche Weise die Stufung der *Ratschlagsabweichungen* (within-subject) erfolgte.

---

[7] Die Versuchsteilnehmerin gab als Initialschätzung (und Finalschätzung) immer den Wert 0 bzw. gar keinen Wert an, sodass in Folge nur Zufallszahlen ohne systematische Abweichungen zur Initialschätzung als Ratschläge dargeboten wurden. Eine Programmfehler scheint wenig wahrscheinlich, da bei den Initialsicherheiten (und Finalsicherheiten) von dieser Person augenscheinlich sinnvolle bzw. variable Angaben gemacht wurden.

7.2.2 Ablauf

Aufgabeninhalt und Ablauf von Experiment 4 stimmten mit denen der vorangegangenen Experimente überein. Um den spezifischen Ziel- bzw. Fragestellungen von Experiment 4 nachgehen zu können, gab es jedoch auch ein paar Veränderungen gegenüber den vorangegangenen Experimenten. Zum einen wurde die Art und Weise der Distanzmanipulation als eine zusätzliche unabhängige Variable (neben der Ratgeberkompetenz) eingeführt. Das heißt, in dem an dieser Stelle geschilderten Experiment wurde der Abstand zwischen den dargebotenen Ratschlägen entweder wie auch in den vorangegangen Experimenten kontinuierlich oder aber kategorial (nah, moderat entfernt, entfernt, sehr weit entfernt) variiert. Die Bedingung mit kontinuierlich abweichenden Ratschlägen verfügte über geringfügig andere Abweichungen als in den vorangegangenen Experimenten, um eine bessere Vergleichbarkeit mit der kategorialen Bedingung zu gewährleisten. Wurde die Abweichung zwischen Ratschlag und Initialschätzung kontinuierlich manipuliert, kamen Abweichungen zwischen ±5% und ±70 jeweils einmal zur Anwendung. Außerdem wurden noch positive Abweichungen im Rahmen von 75% bis 140% jeweils zweimal angewendet. Die Abstufung der einzelnen Abweichungen erfolgte dabei erneut in 5%-Schritten. Die vier Stufen der kategorialen Bedingung wurden so ausgewählt, dass sie den mittleren Abweichungen aus jeweils sieben aufeinanderfolgenden kontinuierlichen Abweichungen entsprachen. Zum Beispiel betrug die „nahe“ Abweichung der kategorialen Bedingung 20%, wobei es sich wiederum um den Mittelwert der kontinuierlichen Abweichungen von 5% bis 35% handelte. Gemäß dieser Vorgehensweise ergaben sich folgende numerische Abweichungen in der kategorialen Bedingung: 20%, 55%, 90% und 125%, wobei die Abweichungen von 20% und 55% sowohl als positive als auch negative Abweichungen angewendet wurden, um dadurch eine maximale Vergleichbarkeit zwischen den beiden Bedingungen der Distanzmanipulation zu ermöglichen. Die Abweichungen von ±20% und ±55% wurden jeweils siebenmal, die Abweichungen von 90% und 125% jeweils vierzehnmal angewendet. Die Versuchsteilnehmer mussten somit im Laufe des Experiments in beiden Bedingungen der Distanzmanipulation 56 Versuchsdurchgänge durchlaufen, wobei der angeblich echte Ratschlag in diesen beiden Bedingungen durchschnittlich gleich stark von der Initialschätzung der Probanden abwich. Wie auch

in Experiment 2, erfolgte in Experiment 4 eine vollständige Randomisierung der Reihenfolge der angewendeten Ratschlagsabweichungen. Dies erschien notwendig, um Konfundierungen zwischen der Reihenfolge der Abweichungen und der Art und Weise ihrer Manipulation (kontinuierlich oder kategorial) zu vermeiden.

7.3 Ergebnisse und Diskussion

7.3.1 Analyse möglicher Störvariablen

Analog zum Prozedere der vorangegangenen drei Experimente wurden zunächst alle AT-Werte pro Versuchsdurchgang und Person bestimmt. Im Anschluss wurden einzelne AT-Werte außerhalb des Bereichs von -1.5 und +1.5 und sowie die kompletten Versuchsdurchgänge (mit den betreffenden Urteilssicherheiten) bei AT-Werten, die nicht auf Grund eines manipulierten Ratschlags berechnet worden sind, von der Analyse ausgeschlossen. Insgesamt wurden dadurch 535 von 5376 Einzelwerten (ca. 10%) nicht mit in die Auswertung einbezogen. Die AT-Mittelwerte aller Probanden wurden im Anschluss mit ihrem Alter korreliert, wobei sich auch hier kein signifikanter Zusammenhang zeigte, $r(96) = -.18$, $p = .08$, obgleich nicht in so deutlichem Maße wie in den vorangegangenen Experimenten. Bei den Probanden in Experiment 4 schien sich daher eine Tendenz anzudeuten, Ratschläge mit zunehmendem Alter weniger zu gewichten. Bei der AV Urteilssicherheit gab es dagegen eindeutig keine systematischen korrelativen Zusammenhänge derselben zum Alter der Probanden, $r(96) = .008$, $p = .94$. Mögliche Geschlechtseffekte auf die Ratschlagsgewichtung und die Urteilssicherheit wurden ebenfalls untersucht. In Experiment 4 beeinflusste dabei das Geschlecht der Probanden systematisch das Ausmaß der Ratschlagsgewichtung, $F(1, 88) = 5.08$, $p = .027$, $\eta^2 = .06$; und zwar dahingehend, dass weibliche Probanden Ratschläge im Mittel stärker gewichteten ($M = .26$, $SD = .18$) als männliche Probanden ($M = .19$, $SD = .19$). Da aber männliche und weibliche Probanden über die verschiedenen Bedingungen hinweg gleichverteilt waren, χ^2 (3, $N = 96$) = 1.24, $p = .74$, wird der Geschlechtseffekt auf die Ratschlagsgewichtung im Folgenden vernachlässigt. Bezüglich der Urteilssicherheit konnten keine Unterschiede zwischen den Geschlechtern festgestellt werden, $F(1, 88) = 0.67$, $p = .42$, $\eta^2 = .008$. Alle möglichen Zweifach- und Dreifach-Interaktionen des Faktors Geschlecht mit den übrigen UVn von Experiment 4 im Hin-

blick auf beide AVn erlangten ebenfalls keine Signifikanz, alle *F*s < 3.8, alle *p*s > .05, alle η^2s < .04.

7.3.2 Ratschlagsgewichtung

Die Analyse der Ratschlagsgewichtung wurde erneut damit begonnen, die AT-Werte für jeden Versuchsdurchgang bei jeder einzelnen Person zu berechnen. In einem nächsten Schritt wurden die so ermittelten AT-Werte hinsichtlich der ihnen vorausgehenden Ratschlagsabweichungen sortiert. Neben dem Zwischensubjektfaktor Ratgeberkompetenz enthielt die ANOVA in Experiment 4 zusätzlich noch den Zwischensubjektfaktor Distanzmanipulation. Um eventuelle Unterschiede im Hinblick auf die AV zwischen den beiden Ausprägungen dieses Faktors untersuchen zu können, war es nötig, dass die Daten aus beiden Bedingungen in ein Datenset zusammengelegt werden können. Damit dies geschehen konnte, mussten jedoch zuvor die unterschiedlichen Abweichungen, die in den beiden Bedingungen verwendet wurden einander numerisch angeglichen werden. In der kontinuierlichen Bedingung wurden zu diesem Zwecke die AT-Werte der Abweichungen gemittelt, die im Mittel den verwendeten kategorialen Abweichung entsprachen. Am konkreten Beispiel erläutert bedeutet dies: Jede kategoriale Abweichung entspricht dem Mittelwert aus jeweils sieben kontinuierlichen Abweichungen. Wie bereits oben aufgeführt, ist somit die kategoriale Abweichung von 20% der Mittelwert aus den kontinuierlichen Abweichungen von 5% bis 35%. Die Abweichung von 55% entspricht dem Mittelwert der kontinuierlichen Abweichungen von 40% bis 70% usw. Der mittlere AT-Wert, dem eine absolute Abweichung von 20% vorausging wurde demnach in der kategorialen Bedingung wie gewohnt bestimmt, indem die berechneten AT-Werte der 14 Versuchsdurchgänge mit einer (absoluten) Ratschlagsabweichung von 20% gemittelt wurden. In der kontinuierlichen Bedingung wurden dazu die AT-Werte aus den 14 Versuchsdurchgängen mit absoluten Ratschlagsabweichungen zwischen 5% und 35% gemittelt. Die so jeweils gemittelten AT-Werte eines jeden Versuchsteilnehmers der kontinuierlichen Distanzbedingung wurden im Anschluss den Ausprägungen der kategorialen Abweichungen als AV zugeordnet, also quasi post-hoc der kategorialen Bedingung angeglichen. Für jede Ver-

suchsperson resultierten auf diese Weise vier mittlere AT-Werte, jeweils einer für eine (mittlere) absolute Abweichung von 20%, 55%, 90% oder 125%.

Die anschließende 2 (*Kompetenz des Ratgebers:* kompetent vs. weniger kompetent) x 2 (*Distanzmanipulation:* kontinuierlich vs. kategorial) x 4 (*Abweichung des Ratschlags von der Initialschätzung:* 20% (5% - 35%), 55% (40% - 70%), 90% (45% - 105%), 125 % (110% - 140%))-Varianzanalyse mit den Zwischensubjektfaktoren Distanzmanipulation und Ratgeberkompetenz sowie dem Innersubjektfaktor Ratschlagsabweichung zeigt, dass –wie auch in Experiment 3– die vermeintliche Ratgeberkompetenz die Ratschlagsgewichtung systematisch beeinflusst, $F(1, 92) = 23.43$, $p < .001$, $\eta^2 = .20$: Ratschläge, die von einem vermeintlich kompetenten Ratgeber stammten, wurden stärker gewichtet ($M = .32$, $SD = .18$) als Ratschläge eines vermeintlich wenig kompetenten Ratgebers ($M = .17$, $SD = .16$). Demnach wird Hypothese 3 angenommen. Es stellte sich weiterhin heraus, dass die Art und Weise der Manipulation der Entfernung zwischen Initialschätzung und Ratschlag keinen systematischen Einfluss auf die mittlere Ratschlagsgewichtung über alle Distanzen hinweg ausübt, $F(1, 92) = 0.13$, $p = .72$, $\eta^2 = .001$, das heißt, die Ratschlagsgewichtung fiel bei kategorialer ($M = .25$, $SD = .18$) und kontinuierlicher Distanzmanipulation ($M = .24$, $SD = .20$) nahezu identisch aus. Im Hinblick auf die Testung meiner Ausgangshypothesen ergab sich erneut ein signifikanter Innersubjekteffekt für die Ratschlagsabweichung, $F(2.00, 184.18) = 10.64$, $p < .001$, $\eta^2 = .10$. Die Art und Weise der Distanzmanipulation zwischen Initialschätzung und Ratschlag übte dabei keinen systematischen Einfluss auf den Zusammenhang zwischen Ratschlagsdistanz und –gewichtung aus. Die Interaktion zwischen Distanzmanipulation und Ratschlagsabweichung erlangte keine Signifikanz, $F(2.00, 184.18) = 0.82$, $p = .44$, $\eta^2 = .01$. Experiment 4 liefert daher eine erste Bestätigung dafür, dass die in den vorangegangenen Experimenten gefundenen Zusammenhänge zwischen Ratschlagsdistanz und –gewichtung kein alleiniges Produkt der dort verwendeten kontinuierlichen Distanzmanipulation darstellen. Hypothese 7a kann angenommen werden. Des Weiteren bestand weder eine signifikante Zweifach-Interaktion zwischen Ratschlagsentfernung und Ratgeberkompetenz (Hypothese 5) oder der Distanzmanipulation und der Ratgeberkompetenz, noch erlangte die Dreifach-Interaktion aller unabhängigen Variablen Signifikanz, alle Fs < 0.5, alle ps > .67 und alle $\eta^2 < .01$. Der Zusammenhang zwi-

schen Ratschlagsentfernung und seiner Gewichtung wurde erneut durch einen quadratischen Kontrast charakterisiert, $F(1, 92) = 21.67$, $p < .001$, $\eta^2 = .19$. Hypothese 1 konnte daher erneut bestätigt werden. Aber auch der lineare Kontrast erlangte wiederum Signifikanz, $F(1, 92) = 14.62$, $p = .034$, $\eta^2 = .05$, auf Grund der Tatsache, dass die mittlere Gewichtung stark entfernter Ratschläge immernoch höher ausfällt als die von nahen Ratschlägen.

Obgleich die Interaktion zwischen der Art der Distanzmanipulation und der Ratschlagsentfernung nicht signifikant wurde, sind die Ergebnisse des Zusammenhangs zwischen Ratschlagsentfernung und –gewichtung in Abbildung 16 für beide Distanzmanipulationsbedingungen (kategorial und kontinuierlich) getrennt dargestellt. Dies geschah um die Ähnlichkeit des Zusammenhangs zwischen beiden Bedingungen noch einmal zu veranschaulichen.

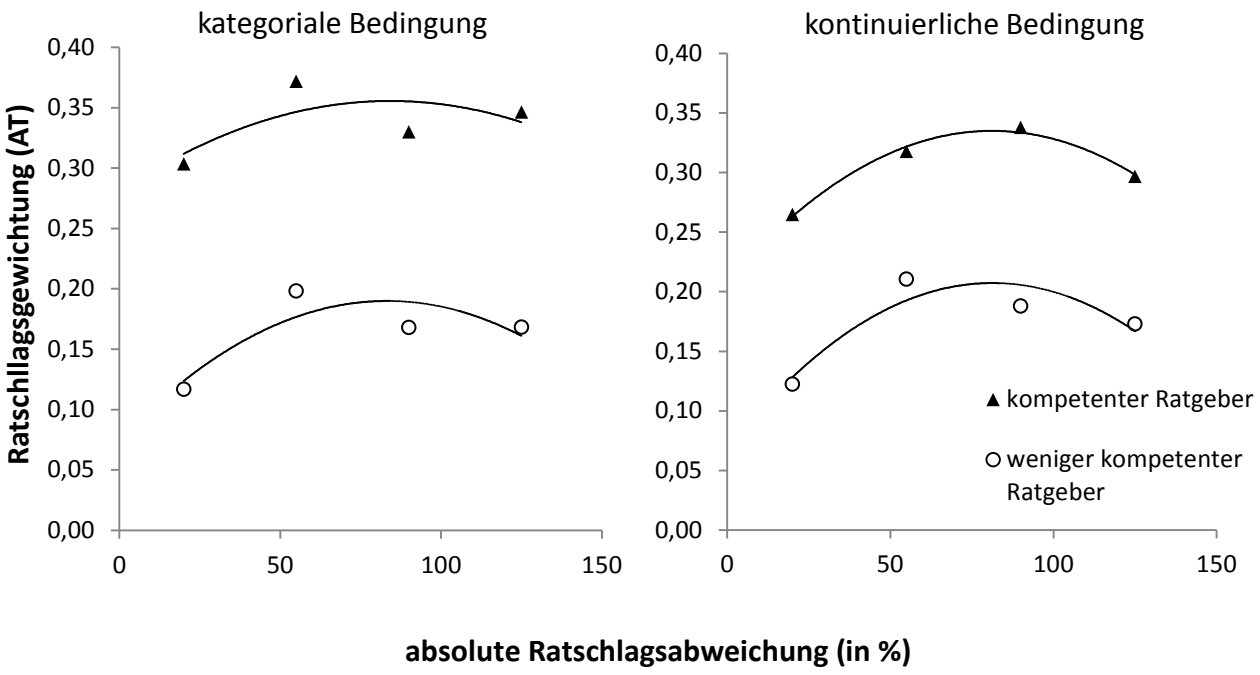

Abbildung 16

grafische Beziehung zwischen Ratschlagsdistanz (in %) und mittlerer Ratschlagsgewichtung für den kompetenten und weniger kompetenten Ratgeber, getrennt dargestellt für beide Distanzbedingungen in Experiment 4

Die anschließende Analyse der Gewichtungsverteilung zeigte ähnliche Ergebnisse wie die in Experiment 3. Insgesamt zeigte sich für die Bedingung mit einem vermeintlich kompetenten Ratgeber eine relativ zwei-gipflige Verteilung der Gewichtungshäufigkeiten, wobei der zweite Gipfel durch zwei unterschiedliche Gewichtungsstrategien geprägt war (siehe Abbildung 17). Bei der Bedingung mit einem weniger kompetenten Ratgeber zeigte sich wie auch in Experiment 3 eine rechtsschiefe Verteilung, welche die Abnahme der Auftretenshäufigkeiten mit zunehmender Höhe der Gewichtung anzeigt. Auf Grund der hiesigen mehrheitlichen Verwendung der Nullgewichtungsstrategie sowie von Gewichtungen in geringem Umfang (bis knapp 40%) ergab sich der Haupteffekt der Ratgeberkompetenz, das heißt im Mittel eine insgesamt niedrigere Gewichtung der Ratschläge eines vermeintlich weniger kompetenten Ratgebers gegenüber denen des kompetenteren.

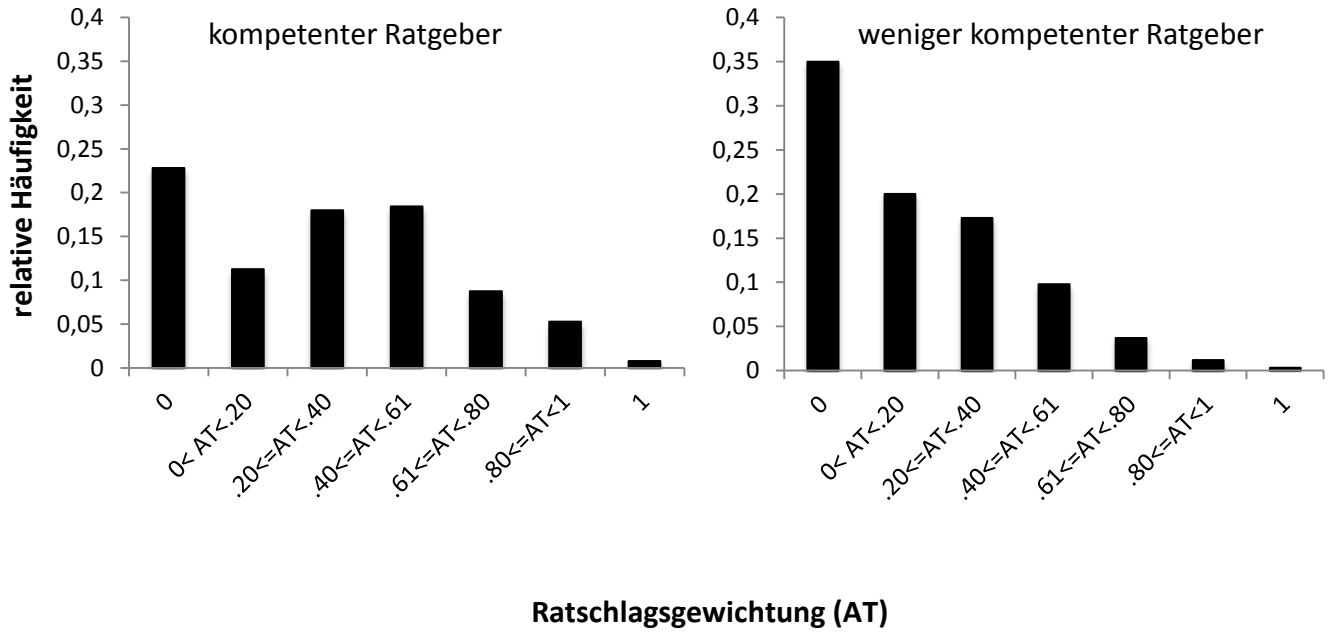

Abbildung 17

Verteilung der Ratschlagsgewichtungen in Experiment 4, getrennt dargestellt für den kompetenten und den weniger kompetenten Ratgeber

Aufgeschlüsselt nach den einzelnen Ratschlagsdistanzen (siehe Abbildung 18), ergab sich für die Häufigkeiten, dass sehr nahe Ratschläge mit sehr großer Häufigkeit komplett ignoriert wurden. In deutlich weniger Fällen gewichteten die Probanden solche Ratschläge geringfügig oder bildeten einen Mittelwert in Kombination mit der ei-

genen Schätzung. Durch den überwiegenden Anteil von Nullgewichtungen resultierte jedoch im Mittel bei nahen Ratschlägen (20% Abweichung) ein recht niedriger AT-Mittelwert. Mit wachsenden Entfernungen nahm der Anteil der Nullgewichtungen jedoch ab, bei gleichzeitiger Zunahme von Gewichtungen bis knapp 40%, und der mittlere AT-Wert dadurch zu. Bei der höchsten Ratschlagsdistanz wurden größtenteils Gewichtungen bis knapp unter 20% vorgenommen. Da der Anteil von Nullgewichtungen hier deutlich geringer ausfiel als bei nahen Ratschlägen resultierte hier im Mittel ein leicht höherer AT-Wert als bei nahen Ratschlägen.

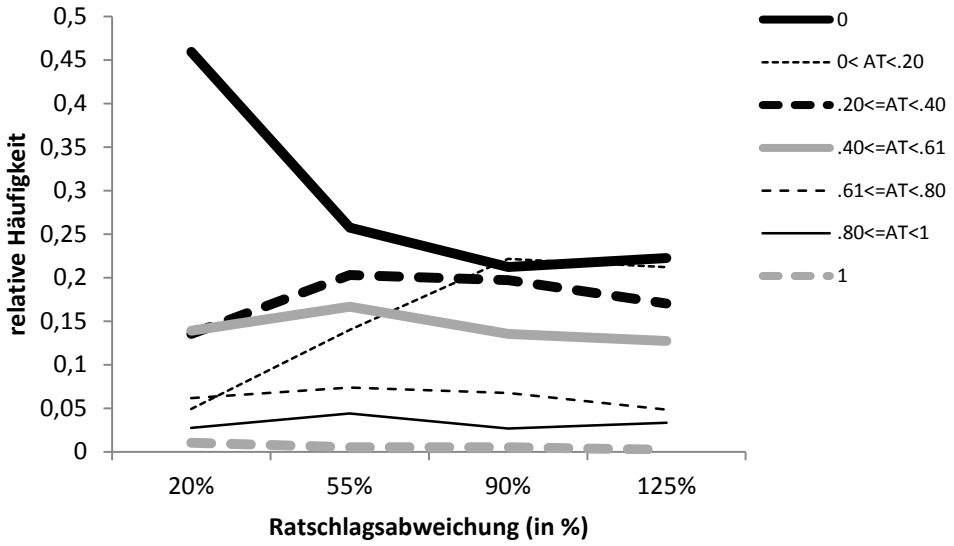

Abbildung 18

relative Häufigkeiten der Ratschlagsgewichtungen in Abhängigkeit der absoluten Ratschlagsentfernung in Experiment 4

7.3.3 Urteilssicherheit

Die Differenz zwischen finaler und initialer Urteilsicherheit pro Person und Versuchsdurchgang wurde im ersten Analyseschritt berechnet. Wie bereits bei der Ratschlagsgewichtung wurden beide Zwischensubjektfaktoren sowie der Innersubjektfaktor in ein Datenset integriert, so dass die Differenzen der Urteilssicherheiten aus der kontinuierlichen Bedingung denjenigen Differenzen der kategorialen Bedingungen zugeordnet werden mussten, deren Werte den Mittelwerten der entsprechenden kontinuierlichen Abweichungen entsprachen. Auch wenn sich deskriptiv eine höhere Ur-

teilssicherheit bei Beratung durch einen vermeintlich kompetenten Ratgeber (*M* = .32, *SD* = .61) gegenüber einem vermeintlich weniger kompetenten Ratgeber (*M* = .17, *SD* = .47) abzeichnet, verfehlt der Haupteffekt der Ratgeberkompetenz in diesem Experiment die Schwelle der Signifikanz, $F(1, 92) = 1.92$, p = .12, $\eta^2 = .03$. Die Art der Distanzmanipulation hat wie zuvor bei der Ratschlagsgewichtung ebenfalls keinen systematischen Einfluss auf die Veränderung der Urteilssicherheit, $F(1, 92) = 0.005$, $p = .94$, $\eta^2 < .001$. Verdeutlicht wird dies durch nahezu identische mittlere Urteilssicherheiten in der kategorialen (*M* = .25, *SD* = .53) und der kontinuierlichen Distanzbedingung (*M* = .25, *SD* = .58). Zusätzlich ergab sich eine signifikante Interaktion der beiden Zwischensubjektfaktoren Ratgeberkompetenz und Distanzmanipulation, $F(1, 92) = 5.77$, $p = .018$, $\eta^2 = .06$. Bei ihrer näheren Betrachtung zeigt sich, dass der höhere Zuwachs der Urteilssicherheit bei Beratung durch einen vermeintlich kompetenten Ratgeber sich lediglich in der kategorialen Ratschlagsbedingung zeigt, $M_{komp} = .43$, *SD* = .50 versus $M_{\neg komp} = .06$, *SD* = .49, wohingegen sich in der kontinuierlichen Bedingung für die beiden Stufen der Ratgeberkompetenz keine unterschiedlichen bzw. deskriptiv sogar eher gegenläufige mittlere Urteilssicherheiten ergeben, $M_{komp} = .21$, *SD* = .69 vs. $M_{\neg komp} = .29$, *SD* = .42. Dementsprechend erlangt der Haupteffekt der Ratgeberkompetenz nur in der kategorialen, $F(1, 44) = 10.796$, $p = .002$, $\eta^2 = .197$, nicht aber in der kontinuierlichen Bedingung Signifikanz, $F(1, 48) = 0.28$, $p = .60$, $\eta^2 = .006$. Hypothese 4 kann daher nur bedingt angenommen werden.

Wie schon in den vorangegangenen Experimenten gab es auch im Experiment 4 einen Distanzeffekt des Ratschlags auf die Urteilssicherheit, $F(2.27, 208.90) = 33.75$, $p < .001$, $\eta^2 = .27$. Zusätzlich zeigte sich hier jedoch eine signifikante Interaktion zwischen der Ratschlagsdistanz und der Ratgeberkompetenz, $F(2.27, 208.90) = 6.07$, $p = .002$, $\eta^2 = .06$. Bei separater Betrachtung der einzelnen UV-Stufen der Ratgeberkompetenz zeigte sich sowohl in der kompetenten Ratgeberbedingung, $F(2.46, 115.73) = 27.63$, $p < .001$, $\eta^2 = .37$, als auch in der weniger kompetenten Ratgeberbedingung ein signifikanter Distanzeffekt, $F(1.92, 86.38) = 7.798$, $p = .001$, $\eta^2 = .15$, sprich, eine von der Entfernung zwischen Initialschätzung und Ratschlag abhängige Urteilssicherheit. Nichtsdestotrotz unterschieden sich die beiden Bedingungen hinsichtlich der Form dieser Beziehung. Wenn Ratschläge durch einen vermeintlich kompetenten Ratgeber erteilt wur-

den, erlangten sowohl der lineare als auch der quadratische Kontrast Signifikanz, $F(1, 47) = 44.14$, $p < .001$, $\eta^2 = .48$ bzw. $F(1, 47) = 15.36$, $p < .001$, $\eta^2 = .25$. Wenn die Ratschläge hingegen von einer angeblich weniger kompetenten Quelle stammten, zeigte sich lediglich ein signifikanter linearer Kontrast, $F(1, 45) = 11.63$, $p = .001$, $\eta^2 = .21$, während der quadratische schwächer ausgeprägt war und daher die Signifikanzschranke verfehlte, $F(1, 45) = 2.88$, $p = .097$, $\eta^2 = .06$ (siehe Abbildung 19). Hypothese 2 kann daher angenommen, und Hypothese 6 muss abgelehnt werden. Eine signifikante Interaktion zwischen Ratschlagsdistanz und der Art der Distanzmanipulation gab es jedoch nicht, $F(2.27, 208.90) = 0.68$, $p = .53$, $\eta^2 = .01$, das heißt die Art und Weise wie der Abstand zwischen Initialschätzung und Ratschlag manipuliert wurde, hatte keinen Einfluss auf die Beziehung zwischen Ratschlagsentfernung und –gewichtung. Hypothese 8a konnte somit angenommen werden.

Alle weiteren Effekte, wie etwa die Dreifach-Interaktion aller unabhängigen Variablen in der Gesamtanalyse oder aber der Haupteffekt Distanzmanipulation sowie mögliche Zweifach-Interaktionen zwischen Ratschlagsdistanz und Distanzmanipulation in den jeweiligen Kompetenzbedingungen erlangten keine Signifikanz, alle Fs < 3.50, alle ps > .07 und alle η^2s < .08.

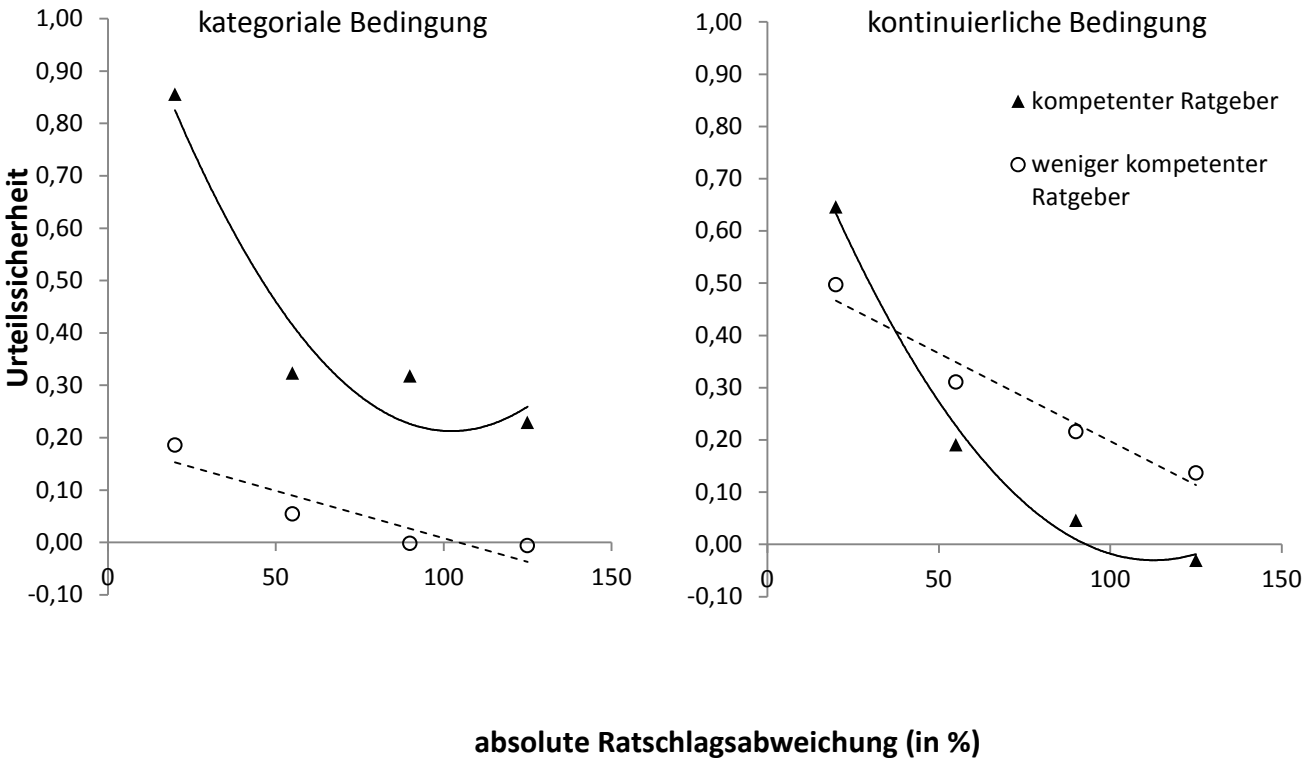

Abbildung 19

grafische Beziehung zwischen Ratschlagsdistanz (in %) und mittlerer Veränderung der Urteilssicherheit für den kompetenten und weniger kompetenten Ratgeber, getrennt dargestellt für beide Distanzbedingungen

Erneut verhielten sich der Verläufe der Mittelwerte der Ratschlagsgewichtung und Urteilssicherheit in allen vier Experimentalbedingungen konträr zueinander. Die Berechnung korrelativer Zusammenhänge zwischen diesen Werten besitzt allerdings wenig Aussagekraft, da vermutlich auf Grund der geringen Anzahl von Mittelwerten pro Bedingung keine Signifikanz erlangt werden konnte. Auch auf Ebene der einzelnen Versuchspersonen wurden die Korrelationen zwischen den einzelnen AT- und Urteilssicherheits-Werten pro Versuchsdurchgang bestimmt – die mittleren Korrelationen der vier Versuchsbedingungen lagen zwischen .02 und .08. Die Anzahl positiver und negativer intraindividueller Korrelationen zwischen Ratschlagsgewichtung und Urteilssicherheit unterschied sich innerhalb und zwischen den Bedingungen nicht signifikant, $\chi^2(3, N = 95) = 1.68$, $p = .64$. Auch hier fallen daher möglicherweise abnehmende Urteilssicherheiten bei stärkerer Ratschlagsewichtung numerisch gesehen stärker ins Gewicht als entsprechende Zunahmen der Urteilssicherheit.

7.3.4 Explorative Datenanalyse: Schätzungsakkuratheit der Judges

Auch in Experiment 4 wurden abschließend sowohl die subjektive als auch die objektive Schätzungsakkuratheit sowie deren Zusammenhang miteinander und anderen Variablen explorativ untersucht. Im Mittel neigten die Probanden auch hier zu einer deutlichen Überschätzung bei quantitativen Urteilen zur Entfernung zwischen zwei Städten, und zwar im Ausmaß von etwa 80% (M_{MPE} = .789, *SD* = .94), $t(95) = 8.296$, $p < .001$, $d = .85$. Im Gegensatz zu den Experimenten 2 und 3 wurde die Korrelation zwischen der eigenen Rangplatzierung unter 100 (vorgeblichen) Teilnehmern und dem MAPE-Wert der Initialschätzungen in Experiment 4 signifikant, $r(91) = .26$, $p = .013$. Betragsmäßig unterscheidet sie sich jedoch nur zu der aus Experiment 2 ($r = .09$), während in Experiment 3 ein ganz ähnlicher, wenngleich nicht signifikanter, Wert erzielt wurde ($r = .26$). Möglicherweise fiel es den Probanden von Experiment 4 also leichter ihre Kompetenz einzuschätzen, als in den Experimenten zuvor, aber mit Sicherheit lässt sich dies nicht aus dem Ergebnis schlussfolgern. Auch die Einschätzung des eigenen Rangplatzes und die Ratschlagsgewichtung weisen einen signifikanten korrelativen Zusammenhang auf, $r(91) = .40$, $p < .001$. Das heißt, je weniger kompetent sich Probanden einschätzten, desto stärker gewichteten sie erhaltene Ratschläge. Einen systematischen Zusammenhang zwischen der eigenen Kompetenzeinschätzung und der durchschnittlichen Veränderung der Urteilssicherheit gab es nicht, $r(91) = .07$, $p = .51$. Im Mittel sahen die Probanden sich und ihre Schätzleistungen auf einem 51. Rang (*M* = 51.31, *SD* = 24.83). Für einen möglicherweise durch die Kompetenzmanipulation hinsichtlich des vermeintlichen Ratgebers ausgelösten Anker-Effekt, so wie er sich in Experiment 3 andeutet, gab es in Experiment 4 keine Hinweise, $t(89) = -.86$, $p = .39$. Personen, die Ratschläge eines vermeintlich kompetenten Ratgebers erhalten, schätzen sich demnach hier nicht signifikant kompetenter ein (*M* = 49.19, *SD* = 25.73) als Probanden mit Ratschlägen eines angeblich weniger kompetenten Ratgebers (*M* = 53.67, *SD* = 23.87).

Werden die Probanden per Mediansplit an Hand des mittleren MAPE-Wertes ihrer Initialschätzungen in zwei Gruppen mit objektiv hoher und niedriger Schätzungsakkuratheit aufgeteilt, und wird anschließend diese Variable als zusätzlicher Zwischen-

subjektfaktor *Schätzungsakkuratheit* (hoch vs. niedrig) in die Haupt-ANOVA integriert, zeigt sich für die AV Ratschlagsgewichtung, wie bereits in Experiment 1 und 2, ein signifikanter Haupteffekt für die Akkuratheit, $F(1, 88) = 20.47$, $p < .001$, $\eta^2 = .19$. Die Gruppe mit einer hohen Schätzungsakkuratheit zeigte hierbei erneut eine signifikant niedrigere mittlere Ratschlagsgewichtung ($M = .18$, $SD = .16$) als die Gruppe mit einer geringeren Schätzungsakkuratheit ($M = .31$, $SD = .19$). Außerdem zeigte sich, dass eine Interaktion zwischen Ratschlagsdistanz und Schätzungsakkuratheit nur knapp nicht signifikant wird, $F(2.03, 178.43) = 2.95$, $p = .054$, $\eta^2 = .03$. In Folge wurden die beiden Akkuratheitsbedingungen daher dennoch separat analysiert. Hierbei stellte sich dann heraus, dass der Distanzeffekt ebenfalls wie in Experiment 2 nur bei den weniger akkuraten Versuchspersonen Signifikanz erlangt, $F(2.22, 97.58) = 10.44$, $p < .001$, $\eta^2 = .19$, nicht aber bei den akkuraten Versuchspersonen, $F(1.74, 76.67) = 2.23$, $p = .12$, $\eta^2 = .05$. In der Bedingung der weniger akkuraten Versuchspersonen wurde der Distanzeffekt zudem sowohl durch einen linearen, $F(1, 44) = 7.83$, $p = .008$, $\eta^2 = .15$, als auch einen quadratischen Kontrast erklärt, $F(1, 44) = 15.62$, $p < .001$, $\eta^2 = .26$. In Abbildung 20 zeigt sich dabei erneut die niedrige Gewichtung der nahen (im Mittel mit 20% abweichenden) Ratschläge gegenüber denen mit einer höheren Abweichung. Für Probanden mit weniger akkuraten Initialschätzungen fiel dieser Gewichtungsunterschied allerdings sichtbar deutlicher aus (.25 bei nahen und .34 bei den entfernteren Ratschlägen), als für akkurater schätzende Probanden (.16 bei nahen und .18 bei entfernteren Ratschlägen).

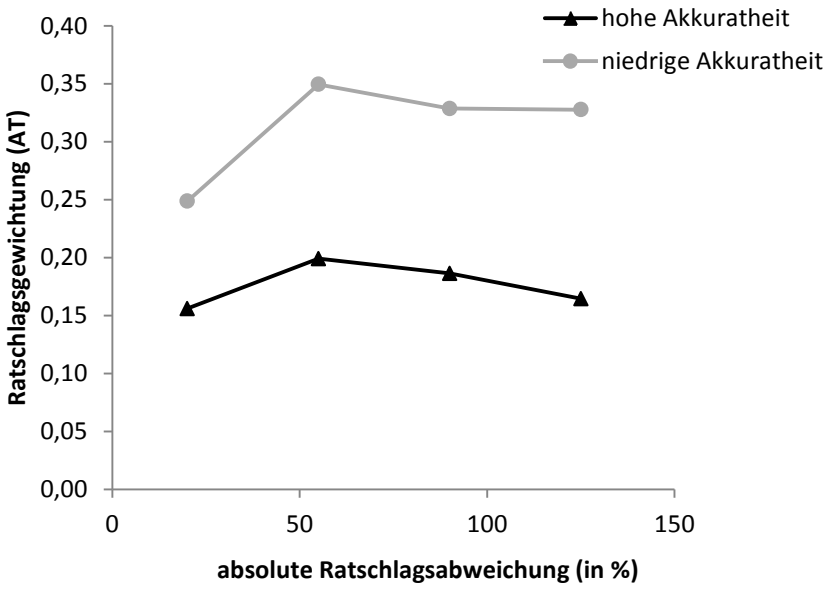

Abbildung 20

grafische Beziehung zwischen Ratschlagsdistanz (in %) und mittlerer Ratschlagsgewichtung in Abhängigkeit von der Schätzungsakkuratheit der Probanden in Experiment 4

Im Hinblick auf die Urteilssicherheit zeigte sich, wie auch in Experiment 1 und 3, kein signifikanter Effekt der Schätzungsakkuratheit auf die mittlere Veränderung der Urteilssicherheit, $F(1, 88) = 0.57$, $p = .45$, $\eta^2 = .01$. So ist der mittlere Sicherheitszuwachs, den weniger akkurate Versuchspersonen durch Ratschläge erhalten, nur unwesentlich höher ($M = .28$, $SD = .64$) als der Sicherheitszuwachs bei akkurateren Versuchspersonen ($M = .21$, $SD = .46$). Eine signifikante Interaktion zwischen der Ratschlagsdistanz und der Schätzungsakkuratheit trat ebenfalls nicht auf, $F(2.24, 196.17) = 0.42$, $p = .68$, $\eta^2 = .005$.

Zusammenfassend zeigen die dargestellten Ergebnisse also, dass unabhängig davon, ob Ratschläge von einer kompetenten oder weniger kompetenten Quelle stammen oder aber ob die Ratschlagsabweichungen kontinuierlich oder deutlich gestufter, sprich kategorial, manipuliert wurden, die Beziehung zwischen Ratschlagsentfernung und –gewichtung stets einem umgekehrten U ähnelt. Damit bestätigen diese Ergebnisse ein weiteres Mal die postulierte Annahme eines quadratischen Zusammenhangs zwischen den beiden Variablen und zeigen weiterhin, dass die durchgehend von Yaniv (2004b) abweichenden Ergebnisse keine Frage der verwendeten Methodik dar-

stellen. Auch hinsichtlich der Gewichtungsverteilungen zeigten sich in Experiment 4 ähnliche Ergebnisse wie in Experiment 3. So wurden nahe Ratschläge (20% Abweichung) überwiegend vollständig diskontiert oder mit der eigenen Initialschätzung gemittelt. Mit zunehmenden Distanzen gewann das Vorgehen der Mittelung an Häufigkeit gegenüber der Null-Gewichtung, aber auch Gewichtungen im Ausmaß des „token Amounts“ sensu Harvey und Fischer (1997) traten nicht selten auf. In der Gesamthäufigkeit kamen sie sogar an zweiter Stelle hinter der Null-Gewichtung, noch vor der Strategie der Bildung eines Mittelwertes. In der Bedingung mit einem vermeintlich weniger kompetenten Ratgeber wurde überwiegend diskontiert, und das über alle manipulierten Ratschlagsdistanzen hinweg. Sofern die erhaltenen Ratschläge gewichtet wurden, geschah dies überwiegend im Rahmen bis 20%.

Die Ergebnisse zur Urteilssicherheit in Experiment 4 gleichen denen der vorangegangen Experiment allerdings nur teilweise. Auf der einen Seite unterstützen sie weiterhin meine Annahme, dass die Entfernung zwischen Initialschätzung und Ratschlag die Urteilssicherheit der Judges systematisch beeinflusst. Auch scheint die Art und Weise der Distanzmanipulation wie schon bei der Ratschlagsgewichtung keinen Einfluss auf die mittlere subjektive Sicherheit oder die Form ihrer Abhängigkeit von der Ratschlagsentfernung auszuüben. Auf der anderen Seite schien diese Urteilssicherheit aber in diesem Experiment in recht inkonsistenter Weise durch die vermeintliche Ratgeberkompetenz beeinflusst zu sein.

Die aufgabenbezogene Kompetenz der Versuchspersonen wirkte sich erneut auf das allgemeine Niveau der Ratschlagsgewichtung aus und es gab wiederum Anzeichen dafür, dass lediglich weniger kompetente Versuchspersonen eine distanzabhängige Gewichtung aufweisen. Die Urteilssicherheit hingegen zeigte sich neuerlich unbeeinflusst von den Fähigkeiten der Probanden.

8. Übergreifende Diskussion

In vier Experimenten wurde untersucht, wie die Entfernung zwischen Initialurteilen von Judges und erhaltenen Ratschlägen die Ratschlagsnutzung und die Urteilssicherheit beeinflusst. Über den bisherigen Forschungsstand hinausgehend wurde postuliert, dass Ratschläge nicht nur in einem geringeren Ausmaß gewichtet werden, wenn sie weit entfernt von der Anfangsschätzung eines Judges liegen, sondern auch, wenn sie dieser recht ähneln bzw. nah an dieser liegen. Weiterhin wurde argumentiert, dass sich zwei qualitativ unterschiedliche Prozesse dafür verantwortlich zeichnen, dass nahe und entfernte Ratschläge in einem vergleichsweise geringen Ausmaß gewichtet werden. Während ich mit der bisherigen Forschung dahingehend übereinstimmte, dass entfernte Ratschläge weniger stark gewichtet werden sollten, weil sie von den Entscheidern als sehr verschieden zu ihrer eigenen Schätzung wahrgenommen werden und daher möglicherweise als wenig plausibel erscheinen (Yaniv, 2004b), stellte ich die Behauptung auf, dass auch sehr nahe Ratschläge wenig gewichtet werden sollten, weil Judges sich durch sie in hohem Maße sozial validiert, das heißt bestätigt fühlen und dadurch subjektiv keine Notwendigkeit mehr besteht, die eigene Ausgangsmeinung in Richtung des Ratschlags zu verschieben. Daraus folgte, dass nahe Ratschläge trotz vorhandener objektiver numerischer Unterschiedlichkeit zur Initialschätzung eine Adjustierung in ihre Richtung überflüssig machen sollten, da sie stattdessen den Entscheidenden in seiner Ausgangmeinung sozial validieren und dadurch mehr bestärken als Ratschläge mit einer moderaten oder gar weiten Entfernung.

Im Rahmen eines Judge-Advisor-Systems waren Probanden dazu angehalten, quantitative Urteilsaufgaben zu bearbeiten. Alle Probanden hatten dabei die Rolle des Judges inne. Die Ergebnisse der vier Experimente erbringen empirische Unterstützung für meine Hypothesen. Ratschläge wurden demnach weniger gewichtet, wenn sie entweder sehr nah oder sehr weit entfernt von der Initialschätzung ausfielen. Gleichzeitig führten nahe Ratschläge zu einem stärkeren Anstieg der Urteilssicherheit der Judges im Hinblick auf die Akkuratheit ihrer Schätzungen, was auf eine Validierungswirkung naher Ratschläge schließen lässt.

8.1 Einfluss der Ratschlagsdistanz auf die Ratschlagsgewichtung

Die vier Experimente meiner Dissertation zeichnen ein eindeutiges Bild des Zusammenhangs zwischen Ratschlagsentfernung und Ratschlagsgewichtung. Gemäß den von mir postulierten Hypothesen ist dieser Zusammenhang umgekehrt U-förmig und spiegelt dabei die besondere Rolle naher Ratschläge bei einem Entscheidungs- oder Urteilsprozess wider. Der gefundene Zusammenhang erweist sich zudem als recht stabil und unabhängig von der vermeintlichen Kompetenz eines Ratgebers oder aber davon, ob die Entfernung zwischen Initialschätzung und Ratschlag eines Probanden kontinuierlich oder in recht groben Unterteilungen manipuliert wurde. An Hand der Unabhängigkeit des Distanzeffektes von der Ratgeberkompetenz manifestiert sich der besondere und zusätzliche Stellenwert, welcher der Ratschlagsdistanz zur Beurteilung einzelner erhaltener Ratschläge zukommt.

Wie bereits eingangs erwähnt, stellt die Ratgeberkompetenz logischerweise einen der stärksten, wenn nicht gar *den* stärksten, Einflussfaktor zur Beurteilung der Güte eines erhaltenen Ratschlages dar (Bonaccio & Dalal, 2006; Harvey & Fischer, 1997). Man könnte also meinen, dass eine so grundlegende Information wie die über die Expertise bzw. Kompetenz eines Ratgebers den meisten Menschen genügen sollte, um den erhaltenen Ratschlägen in einem gewissen Ausmaß Folge zu leisten. Mit Blick auf die mittlere Gewichtung der Ratschläge unterschiedlich kompetenter Ratgeber trifft dies auch zu – so zeigt sich in den Experimenten 3 und 4, in denen die vermeintliche Ratgeberkompetenz variiert wurde, stets eine höhere durchschnittliche Gewichtung der Empfehlungen eines vermeintlich kompetenten Ratgebers gegenüber denen eines weniger kompetenten. Trotzdem lassen die entsprechenden Ergebnisgraphiken sowie ausbleibende Interaktionen zwischen den Variablen Distanz und Kompetenz darauf schließen, dass sich innerhalb jeder Kompetenzbedingung der postulierte umgekehrt U-förmige Distanzeffekt zeigt. Dies lässt wiederum den Schluss zu, dass zur Beurteilung der grundsätzlichen Güte der Ratschläge die Ratgeberkompetenz die ausschlaggebende Information darstellt. Im Einzelfall wird jedoch anscheinend auch immer die automatisch mit dem Ratschlag einhergehende bzw. „mitgelieferte“ Information seines Abstandes von der eigenen Initialschätzung ausgewertet und die Gewichtung in Folge

dementsprechend angepasst. Eventuell besitzt die Ratschlagsdistanz also einen inkrementellen informationalen Wert hinsichtlich der Qualität der vom Ratgeber stammenden Empfehlungen. Vergangene Forschung hat bereits gezeigt, dass Judges ihren Ratgebern, von denen sie zu mehreren Gelegenheiten wiederholt Ratschläge erhalten, keine festen Gewichte zuordnen, sondern diese in einem Rahmen von 0% bis 100% variieren, höchstwahrscheinlich in der Absicht, die Akkuratheit ihrer Finalschätzungen zu erhöhen (Soll & Larrick, 2009; Soll & Mannes, 2011). Veränderungen in der Ratschlagsdistanz mögen daher einen der Faktoren darstellen, denen die Variabilität der Ratschlagsgewichtung durch die Judges zu Grunde liegt, wofür nicht zuletzt auch die Ergebnisse der Analysen der Gewichtungsverteilungen sprechen.

Die Nutzung von Ratschlagsdistanz als eine Art Qualitätsindikator kann aber gegenwärtig nur gemutmaßt werden. Dieser Qualitätsaspekt besitzt dabei genau genommen zwei Komponenten – eine objektive und eine subjektive. Um Aussagen über die objektiven Zusammenhänge zwischen Ratschlagsdistanz und –qualität treffen zu können, müssten Untersuchungen unternommen werden, in denen echte Ratschläge präsentiert werden und der Abstand zwischen Initialschätzung und diesen Ratschlägen nicht in festgelegter Weise manipuliert wird, sondern natürlich vorkommenden Schwankungen unterliegt. Eine weitere Möglichkeit wäre auch damit gegeben, bei vorhandenen Datensätzen eine Person als Judge zu fixieren und die anderen und deren Schätzungen (künstlich) als Ratgeber bzw. Ratschläge zu behandeln. Zwischen den Initialschätzungen des fixierten Judges und den Schätzungen der als Ratgeber fungierenden Personen könnten zum einen Entfernungsmaße (absolut und prozentual) und zum anderen Akkuratheitswerte für diese künstlichen Ratschläge bestimmt und miteinander korreliert werden, um Aussagen über die Stärke des Zusammenhangs zwischen beiden Größen treffen zu können. Zusätzlich wären Analysen hinsichtlich eines eventuell moderierenden Effekts der Judge- und/oder Ratgeberkompetenz möglich.

In einem weiteren Schritt ließen sich etwa durch die Ermittlung von Zusammenhängen zwischen der Ratschlagsdistanz, Gewichtung und der (Veränderung der) Qualität der Finalschätzungen auch noch Aussagen darüber treffen, ob Probanden im Mittel zu besseren Finalschätzungen gelangen, wenn sie Ratschläge distanzabhängig

gewichten oder aber ein relativ gleichbleibendes Gewichtungsniveau aufweisen. Um festzustellen, ob die Ratschlagsentfernung tatsächlich die wahrgenommene, also, subjektive, Qualität beeinflusst, könnte in zukünftigen Untersuchungen zusätzlich noch die wahrgenommene Plausibilität und/oder Qualität der dargebotenen, in ihrer Distanz zur Initialschätzung systematisch abweichenden Ratschläge miterhoben werden.

Meine Ergebnisse widersprechen zum Teil den von Yaniv (2004b) ermittelten bzw. ergänzen diese um die Erkenntnisse zur Gewichtung besonders naher Ratschläge. Wie eingangs erwähnt, beinhalten Yanivs Ergebnisse, dass Ratschläge umso weniger gewichtet wurden, je weiter entfernt sie von der Initialschätzung waren. Während die Signifikanz des quadratischen Kontrasts einerseits darauf schließen lässt, dass auch im Rahmen meiner Experimente abnehmende Gewichtungen bei entfernten Ratschlägen auftraten, bedeutet sie andererseits gleichzeitig, dass sich das gleiche Phänomen zusätzlich auch bei sehr nahen Ratschlägen zeigt. Im Gegensatz zu den Experimenten von Yaniv nimmt also die Gewichtung von Ratschlägen mit zunehmender Distanz zur Initialschätzung im Rahmen meiner Experiment anfangs sogar erst einmal zu, bevor sie dann wieder abnimmt (und zwar im Vergleich zum anfänglichen Anstieg weniger stark), weswegen der lineare Kontrast auch einen steigenden Trend anzeigt.

Meine Experimente schließen dabei Unterschiede in der methodischen Vorgehensweise, wie etwa hinsichtlich der Art der Manipulation des Abstandes zwischen Initialschätzung und Ratschlag, bei der Datenerhebung sowie Spezifika der verwendeten Urteilsaufgaben als Ursache der empirischen Diskrepanz zu Yaniv (2004b) weitestgehend aus. Die naheliegendste Erklärung ist daher, dass nahe Ratschläge aus Yanivs Originalstudie noch immer zu stark von den Initialschätzungen der Probanden abwichen, um einen Assimiliationsprozess in Gang zu setzen. Auf den ersten Blick mag dies etwas unwahrscheinlich erscheinen, wenn man bedenkt, dass die mittleren absoluten Unterschiede (in Jahreszahlen) zwischen den Initialschätzungen der Probanden und den erhaltenen Ratschlägen bei Yaniv überwiegend nicht die Höhe der prozentualen Abweichungswerte der kategorialen Bedingung meines Experiment 4 erreichen dürften. Nichtsdestotrotz kann der Umstand, dass ich eine andere Urteilsaufgabe als Yaniv verwendet habe, dazu geführt haben, dass geringe und hohe absolute (und dadurch

letztlich auch prozentuale) Abweichungen in den zwei verschiedenen Aufgabentypen von den Probanden als recht unterschiedlich wahrgenommen wurden. Demgemäß ist nicht auszuschließen, dass Teilnehmer der Untersuchung von Yaniv einen um 20 Jahre abweichenden Ratschlag als grundlegend verschiedener zu ihrer Initialschätzung empfunden haben, als die Probanden meiner Experimente einen um 20km abweichenden Ratschlag, so dass bei Yaniv so etwas wie ein Kontrast-Effekt möglicherweise recht schnell eintrat (vgl. auch Nemeth & Endicott, 1976).

Die Analysen der Gewichtungsverteilungen über die verschiedenen Ausprägungen der Ratschlagsentfernungen hinweg gaben zusätzlichen Aufschluss über das Zustandekommen des Distanzeffekts. Es zeigte sich, dass die Versuchsteilnehmer weniger das durchschnittliche Ausmaß ihrer Gewichtungen in Abhängigkeit der Ratschlagsdistanz anpassten, sondern die Anwendungshäufigkeit zweier (teilweise auch dreier) Strategien der Urteilsgenerierung variierte. Soll und Larrick (2009) sprachen in ihrer Untersuchung in diesem Zusammenhang von *Averaging* und *Choosing* als diesen beiden bevorzugten Strategien – also, von der Bildung eines Mittelwertes zwischen Inititalschätzung und Ratschlag oder der Wahl eines dieser beiden Urteile. Es zeigte sich, dass auch im Rahmen meiner Experimente häufig gemittelt oder gewählt wurde. Bei der Wahl-Strategie gab es dabei ungeachtet der Ratschlagsdistanz eine deutliche Präferenz zur Wahl bzw. Beibehaltung des eigenen Initialurteils gegenüber der Übernahme des Ratschlags. Die Vollgewichtung von Ratschlägen kam zwar vor, allerdings nur in Einzelfällen. Für den Verlauf der Verwendung dieser Strategien über die Ratschlagsdistanzen hinweg lässt sich vereinfacht sagen, dass bei sehr geringen Ratschlagsentfernungen Ratschläge überwiegend vollständig diskontiert wurden, was die geringen mittleren AT-Werte bei diesen Abweichungen zur Folge hatte. Die Häufigkeitskurve der Nullgewichtung nimmt jedoch mit zunehmenden Ratschlagsdistanzen fast kontinuierlich ab, während andere Gewichtungsstrategien (bis knapp 40% oder die *Averaging*-Strategie) mit zunehmenden Distanzen häufiger vorkommen, so dass es dadurch zu einem Anstieg der mittleren Gewichtung kommt. Bei den höchsten Abweichungen verhält sich diese Tendenz dann wieder rückläufig und der Anteil der Nullgewichtungen nimmt wieder leicht zu. Dennoch kommen sie bei weitem nicht so häufig vor wie bei sehr nahen Ratschlägen und andererseits finden hier zusätzlich noch deutlich mehr Gewich-

tungen in geringerem Umfang (bis 20% oder bis knapp 40%) statt, so dass der AT-Mittelwert zwar wieder sinkt, aber nicht auf das Niveau wie bei sehr nahen Ratschlägen abfällt. Die Folge ist der wiederholt berichtete umgekehrt U-förmige Zusammenhang zwischen Ratschlagsentfernung und –gewichtung.

Die Häufigkeitsverteilungen der Gewichtungen waren weniger W-förmig als bei Soll und Larrick (2009), sondern eher zwei-gipflig, wobei der zweite Gipfel, neben der Nullgewichtung der Ratschläge, in den Experimenten 1 und 2 bei der Strategie des Mittelns lag. Bei der Verwendung höherer Abweichungen wie bei Experiment 3 und 4 lag in der Bedingung mit einem angeblich kompetenten Ratgeber der zweite Gipfel dann aber bei der Kategorie mit Gewichtungen im Rahmen von 20% bis knapp 40%. Dies ist wahrscheinlich auf deren erhöhtes Auftreten bei höheren Ratschlagsdistanzen zurückzuführen, wo den Probanden die Bildung eines Mittelwertes womöglich auf Grund der hohen Diskrepanz nicht mehr angemessen schien und stattdessen das eigene Urteil stärker gewichtet wurde. Die Häufigkeitsverteilung in den Bedingungen mit einem weniger kompetenten Ratgeber war nicht mehr zwei-gipflig, sondern rechtsschief mit immer geringer werdenden Häufigkeiten, je größere Gewichtungen die einzelnen Kategorien anzeigten.

Die bedingungsübergreifend hohe Anzahl von Nullgewichtungen von Ratschlägen bei sehr niedrigen Distanzen unterstützt zusätzlich die Annahme eines Assimilationseffektes. Zwar bestand, wie gesagt, bei den Versuchsteilnehmern eine generelle Präferenz für die Beibehaltung der eigenen Meinung; diese war aber noch deutlicher ausgeprägt, wenn der Ratschlag dem eigenen Initialurteil der Probanden recht ähnlich war (bis ca. 15% Abweichung bzw. auch bei der 20%-Abweichungsstufe in Experiment 4). Dies deutet darauf hin, dass in diesem Diskrepanzbereich numerische Unterschiede zwischen den beiden Schätzungen für die Probanden nicht so ausschlaggebend waren, dass ihnen statt der Beibehaltung der eigenen Annahme beispielsweise die Bildung eines Mittelwertes aus beiden Schätzungen angemessener erschien. Die Ähnlichkeit wurde daher vermutlich hier häufig überschätzt.

8.2 Einfluss der Ratschlagsdistanz auf die Urteilssicherheit

Darüber, ob mit der Integration eines Ratschlags in ein Finalurteil das Ziel der größtmöglichen Urteils- oder Entscheidungsqualität erreicht werden konnte, kann häufig nicht unmittelbar im Anschluss an den Entscheidungsprozess (z.B. durch unmittelbares Feedback) Auskunft erlangt werden. Die objektive Güte von getroffenen Entscheidungen oder Urteilen durch Konsultation einer oder mehrerer Ratgeber offenbart sich in den meisten Alltagssituationen erst mit einiger Verzögerung. Daher wird die Entscheidungsqualität in der Regel zunächst (und zumeist automatisch) ganz subjektiv beurteilt. Wie gut oder schlecht eine getroffene Entscheidung oder ein getroffenes Urteil subjektiv von Judge beurteilt werden, drückt sich in dessen Urteilssicherheit aus (Sniezek & Buckley, 1995). Im Rahmen meiner Dissertation wollte ich nun erste systematische Erkenntnisse darüber gewinnen, wie sich die Urteilssicherheit in Abhängigkeit der Ratschlagsentfernung verhält. In Übereinstimmung mit eingangs erwähnten Forschungsergebnissen (Heath & Gonzales, 1995; siehe auch Kapitel 3.1) konnte im Rahmen der von mir durchgeführten Experimente bestätigt werden, dass die subjektive Urteilssicherheit im Laufe eines Beratungsdurchganges im Mittel grundsätzlich zunimmt. Diese Zunahme erweist sich als teilweise abhängig von der Beraterkompetenz; und zwar auf die Art, dass sie bei Beratung durch einen vermeintlich kompetenten Ratgeber im Mittel höher ausfällt als bei einem angeblich weniger kompetenten. Zumindest konnte dies in Experiment 3 eindeutig gezeigt werden, in Experiment 4 zeigt sich dieser Effekt nur deskriptiv.

Auch für die AV Urteilssicherheit zeigt sich zudem, dass unabhängig vom Kompetenzeffekt auch die mit dem Ratschlag einhergehende Information seines Abstandes zur eigenen Initialschätzung wahrgenommen wird. Diese Wahrnehmung manifestiert sich dabei in systematischen und replizierbaren Veränderungen der Urteilssicherheit. Bei nahen Ratschlägen ist der Zuwachs somit deutlich ausgeprägter als bei zunehmend entfernteren Ratschlägen. Je mehr also der Ratschlag von der eigenen Anfangsmeinung abweicht, desto weniger Vertrauen in die Qualität bzw. Akkuratheit des Finalurteils scheint zu bestehen.

8.3 Adjustierung versus soziale Validierung

Ein niedrige bis nicht vorhandene Gewichtung naher Ratschläge bei gleichzeitigem Anstieg der subjektiven Urteilssicherheit spricht für die Annahme, dass nahe Ratschläge in diesem Fall nicht missachtet wurden. Stattdessen spiegelt sich ihre Nutzung in einem Anstieg der Urteilssicherheit wider. Der deutliche Anstieg der Urteilssicherheit verbunden mit der geringen Gewichtung bei nahen Ratschlägen spricht dafür, dass Judges sich durch Ratschläge, die nah an ihrer Initialschätzung lagen, sozial validiert fühlten. Im Detail, wenn Ratschläge genügend nah am Initialurteil liegen, führt dies vermutlich dazu, dass Judges beide Angaben als nahezu gleich empfinden und ihnen daher eine Anpassung des Finalurteils an den Ratschlag nicht nötig erscheint, weil sie gleichzeitig sehr geneigt scheinen, ihr eigenes Urteil als das akkuratere von zwei sehr ähnlichen Urteilen anzusehen. Die bereits von Gino und Moore (2007) geäußerte Vermutung, dass eine sehr niedrige Gewichtung des Ratschlags bei hoher Ähnlichkeit desselben zur Initialschätzung eines Judges nicht zwangsläufig keinerlei Ratschlagsnutzung bedeutet, weil der Judge nach der Bestätigung seiner eigenen Schätzung durch den Ratschlag letztere einfach beibehält, kann also als bestätigt angesehen werden.

Meine Ergebnisse legen somit nahe, dass Ratschläge auf mindestens zwei Arten genutzt werden können, nämlich Adjustierung der eigenen Ausgangsmeinung auf der einen und Veränderung der Sicherheit bezüglich der eigenen Meinung auf der anderen Seite. Vergangene Forschung hat sowohl Ratschlagsgewichtung als auch Veränderungen der Urteilssicherheit im JAS untersucht (siehe Bonaccio & Dalal, 2006), aber mit nur wenigen Ausnahmen wie etwa der Leitstudie von Sniezek und Buckley (1995) wurden diese Variablen stets getrennt untersucht. Meine Ergebnisse zeigen, dass die Ratschlagsgewichtung allein nicht das vollständige Ausmaß der tatsächlichen Ratschlagsnutzung wiedergeben kann. Stattdessen scheint die wahrgenommene Ähnlichkeit zwischen Initialschätzungen und Ratschlägen einen eigenen informationalen Charakter zu besitzen, wobei die Nutzung dieser Information sich in einem Anstieg der Urteilssicherheit widerspiegelt. Interessant ist dabei, dass vergangene Untersuchungen somit möglicherweise das tatsächliche Ausmaß der Ratschlagsnutzung unterschätzt haben könnten. So hatte ich ja bereits eingangs erwähnt, dass Judges oftmals Ratschläge

nicht in dem Ausmaß nutzen wie es ihre objektive Qualität nahelegt (Bonaccio & Dalal, 2006; Yaniv, 2004a; Yaniv, 2004b; Kapitel 1 und 2.3). Wenn nun zusätzlich das Ausmaß betrachtet würde, in dem Judges die erhaltenen Ratschläge als Quelle sozialer Validierung nutzen, könnte die Tendenz, Ratschläge zu diskontieren, de facto kleiner ausfallen als ursprünglich angenommen. In diesem Zusammenhang sprechen die hier referierten Daten dafür, dass es durchaus sinnvoll wäre, in zukünftigen Untersuchungen gleichzeitig sowohl das Ausmaß der Urteilsverschiebung in Richtung Ratschlag als auch Veränderungen der Urteilssicherheit der Judges zu erheben, um ein vollständigeres Bild davon zu erhalten, wie bzw. in welchem Umfang Ratschläge genutzt werden.

8.4 Schätzungsakkuratheit der Judges

Da im Rahmen dieser Doktorarbeit keine realen Ratschläge dargeboten wurden, wurde auf die Formulierung spezifischer Hypothesen zur Schätzungsakkuratheit der Judges und eine mögliche Veränderungen derselben durch den Erhalt von Ratschlägen verzichtet. Stattdessen wurde, auch in Anlehnung an die von Yaniv (2004b) berichteten Befunde, explorativ untersucht, welche Zusammenhänge zwischen subjektiver und/oder objektiver Kompetenz- bzw. Akkuratheitseinschätzung der Judges und den abhängigen Variablen, dabei insbesondere der Ratschlagsgewichtung, bestehen.

Insgesamt zeigte sich eine Tendenz zur Überschätzung der wahren Werte. Die Versuchsteilnehmer überschätzten also im Mittel den Kaloriengehalt ausgewählter Lebensmittel sowie die Entfernung zwischen europäischen Hauptstädten. Des Weiteren deuten die Experimente darauf hin, dass es den Probanden recht selten gelingt, ihre Kompetenzen richtig einzuschätzen. Ein signifikanter Zusammenhang zwischen subjektiven und objektiven Kompetenzindikatoren bestand bei nur einem Experiment.

Unabhängig vom Aufgabeninhalt zeigte sich mit Ausnahme von Experiment 3, dass die objektive Schätzungsakkuratheit der Probanden das grundlegende Niveau der Ratschlagsgewichtung beeinflusst, so dass akkurater schätzende Probanden erhaltene Ratschläge im Mittel geringer gewichten als weniger akkurat schätzende Probanden. In Übereinstimmung damit zeigten zwei von drei Experimenten, in denen die subjektive Akkuratheit erhoben wurde, außerdem signifikante korrelative Zusammenhänge zwi-

schen der Einschätzung der versuchspersonen-eigenen Kompetenz und der mittleren Gewichtung der Ratschläge dergestalt, dass mit abnehmender subjektiver Kompetenz die mittlere Gewichtung der Ratschläge zunahm. Dieses Ergebnis ist aber in der Tat ein wenig überraschend, da auf Grund der Zusammenhänge zwischen der subjektiven Bewertung der eigenen Schätzleistung und objektiven Akkuratheitsmaßen der Eindruck entstehen kann, dass die Versuchsteilnehmer die Güte ihrer eigenen Leistungen nicht zutreffend einschätzen können. Auf Grund dieses mangelnden Vermögens kommt eine bewusste Steuerung der Gewichtung erhaltener Ratschläge als Erklärung für den Akkuratheitseffekt sehr wahrscheinlich nicht in Frage, obgleich in allen Experimenten, in denen beide Maße erhoben wurden, schwach positive Zusammenhänge bestehen.

Andererseits besteht die Möglichkeit, dass das von mir gewählte Maß zur Überprüfung der subjektiven Fähigkeit die tatsächlich vermutete Eigenkompetenz nicht fehlerfrei abzubilden vermag. Die Frage nach dem eigenen Rang unter 100 Versuchsteilnehmern war demzufolge möglicherweise zu sehr durch den vorgegebenen Rang des Ratgebers beeinflusst. So ist davon auszugehen, dass die Probanden ihre eigene Leistung stets an der des Ratgebers messen. Wenn dessen angeblicher Rang nun, wie stets geschehen, bekannt gegeben wird, wird der eigene zwangsläufig daran bemessen. Experiment 3 zeigte wohin dies im Extremfall führen kann – nämlich dazu, dass Probanden in der Bedingung mit einem vermeintlich kompetenten Ratgeber sich und ihre Leistungen signifikant besser einschätzen, als Probanden in der Bedingung mit einem angeblich weniger kompetenten Ratgeber. Auch mit Selbstauskünften typischerweise einhergehende Effekte im Rahmen der sozialen Erwünschtheit können die Genauigkeit des Maßes schmälern. Auch wenn sich Probanden unter Umständen berechtigterweise selbst für recht kompetent hielten, würden sie dies vielleicht nicht in der angemessenen Deutlichkeit äußern, um nicht als unbescheiden wahrgenommen zu werden. Umgekehrt neigen wissentlich weniger kompetente Probanden möglicherweise dazu, die Frage nach dem eigenen Kompetenzrang eher selbstwertdienlich zu beantworten. Auf Grund der starken Gewährleistung der Anonymität der Probanden in den Experimenten bleibt fraglich, ob und inwieweit diese Mechanismen tatsächlich wirksam waren. Es ist demnach möglich, dass die Angabe der eigenen Platzierung un-

ter 100 angeblichen Probanden kein hinreichend valides Maß für die subjektive Einschätzung der Judge-Kompetenz darstellt.

Weiterhin deutet sich in den Experimenten an, dass die Schätzungsakkuratheit den gefundenen Distanzeffekt der Ratschlagsgewichtung nur unsystematisch moderiert, und wenn dann dahingehend, dass eine signifikante Abhängigkeit der Ratschlagsgewichtung von der Distanz zwischen Ratschlag und Initialschätzung vor allem dann auftritt, wenn Probanden weniger gute Initialschätzungen abgeben, und nur in geringerem Ausmaß, wenn sie zu den akkurateren Versuchspersonen zählen. In Experiment 2 wurde dies durch eine signifikante Interaktion zwischen Ratschlagsdistanz und Schhätzungsakkuratheit sowie die beschriebenen bedingten Haupteffekte angezeigt. In Experiment 4 verfehlte die Interaktion zwischen Distanz und Schätzungsakkuratheit die Schwelle der Signifikanz minimal. Separate Analysen für die zwei unterschiedlichen Akkuratheitsgruppen zeigten jedoch auch in diesem Experiment an, dass der Distanzeffekt nur für die Gruppe der weniger akkuraten Versuchspersonen Signifikanz erlangt - wenngleich das Ergebnismuster auch für die akkurateren Versuchspersonen ähnlich war, nur eben schwächer ausgeprägt.

Für die Gruppe der weniger akkurat schätzenden Personen werden daher vergleichbare Ergebnisse wie die von Yaniv (2004b) berichteten gefunden, wenngleich die umgekehrte U-Form des Zusammenhangs zwischen Ratschlagsdistanz und -gewichtung von ihm nicht hypostasiert war. Hinsichtlich der Ergebnisse für akkurater schätzende Personen bestehen jedoch Unterschiede. Während in meinen Experimenten diese Personen teilweise erst gar keine distanzabhängige Gewichtung der Ratschläge aufwiesen, zeigt sich in den Ergebnissen von Yaniv (2004b) stets ein signifikanter Distanzeffekt in Form eines monoton fallenden Zusammenhangs zwischen den beiden Variablen. Möglicherweise waren akkurater schätzende Personen im Rahmen meiner Experimente stärker als bei Yaniv (2004b) davon überzeugt, dass Schwankungen der Distanz zwischen eigenen Initialschätzungen und erhaltenen Ratschlägen ganz natürlich und nicht zwangsläufig ein Indikator für die spezifische Ratschlagsqualität sind, weswegen die Gewichtung diesen Schwankungen auch nicht in erheblichen Ausmaß angepasst werden muss.

Die Auswirkungen der Schätzungsakkuratheit auf die Urteilssicherheit fallen eindeutiger aus. Der Haupteffekt wird in nur einem von vier vorhandenen Experimenten signifikant, so dass sich andeutet, dass weniger akkurate Versuchspersonen nicht unbedingt stärker sozial von Ratschlägen profitieren können als leistungsstärkere Versuchspersonen, sondern dass unabhängig von der eigenen Kompetenz alle im (fast) gleichen Ausmaß von erhaltenen Ratschlägen im Sinne einer sozialen Validierung profitieren können.

Zusammenfassend liefern die hier berichteten Ergebnisse erste Hinweise darauf, welche Auswirkungen die judge-eigene Kompetenz auf die Ratschlagsgewichtung und Urteilssicherheit sowie die diesbezüglich auftretenden Distanzeffekte hat. In Zukunft sollten derlei objektive Maße immer Teil der Erhebungen sein, um allgemeingültigere Aussagen über die Schätzleistungen der Versuchsteilnehmer und die Passung jener mit der von ihnen subjektiv eingeschätzten Kompetenz treffen zu können. Nicht zuletzt deshalb, weil Probanden sich anscheinend auch in manchen Fällen durch die vorgegebene Kompetenzmanipulation bei den Ratgebern hinsichtlich der Einschätzung ihrer eigenen Expertise systematisch beeinflussen lassen. Experiment 3 zeigte dies eindrucksvoll, aber an Hand der Tatsache, dass dieser Befund in Experiment 4 nicht signifikant repliziert werden konnte, ist es momentan nicht möglich, Aussagen darüber zu treffen, ob es sich dabei um ein belastbares Phänomen handelt.

8.5 Beschränkungen und Ausblick auf die weitere Forschung

Abschließend soll noch auf die Grenzen meiner Untersuchungen eingegangen werden. An erster Stelle stehen dabei die Inhalte der Schätzaufgaben. Mit den Entfernungs- und Kalorienschätzungen wurde ein relativ kleines Feld des Alltagswissens abgedeckt. Dementsprechend lässt sich nicht gänzlich ausschließen, dass die gefundenen Beziehungen zwischen Ratschlagsentfernung und –gewichtung bzw. der Urteilssicherheit nicht zumindest teilweise Spezifika solcher numerischer Abfragen sind. Zukünftig ist es deshalb unumgänglich zu testen, ob die hier berichteten Ergebnisse sich beispielsweise auf ökologisch validere Urteilsaufgaben wie etwa Finanz- oder Klimaprognosen übertragen lassen.

Zweitens wurden in den Experimenten keine echten Ratschläge (z.B. durch Darbietung von Schätzungen früherer Probanden) verwendet, sondern Ratschläge wurden als relative Abweichungen zu den Initialschätzungen der Probanden generiert. Während dies einerseits eine volle experimentelle Kontrolle über die Manipulation der Ratschlagsentfernung ermöglichte und das Stimulusmaterial über die verschiedenen Experimentalbedingungen vergleichbar machte, ging dies auf Kosten der Schlussfolgerungen, die hinsichtlich Angemessenheit der Gewichtungsstrategien und Veränderung der Urteilssicherheit der Judges gezogen werden konnten. Zum Beispiel ist es so, dass die Anstiege der Urteilssicherheit bei nahen Ratschlägen nur angemessen wären, wenn Konsens überwiegend Korrektheit implizierte. Diese Schlussfolgerung ist jedoch nur möglich im Falle unabhängiger Schätzungen von Judge und Advisor sowie jeweils niedrigen Urteilsverzerrungen (sog. *Biases*) im Sinne einer grundlegenden Tendenz zur Über- oder Unterschätzung der Werte. Die gegenwärtige Forschung deutet darauf hin, dass Judges nicht sensitiv gegenüber solchen Urteilsverzerrungen sind und dass in Folge die Abgabe eines Konsensurteils zu *Overconfidence*, also, einer unangemessen hohen subjektiven Sicherheit, führen kann, da ein solches Urteil tatsächlich keine erhöhte Akkuratheit aufweist (Lorenz, Rauhut, Schweitzer & Helbing, 2011). Die logische Konsequenz dieser Ausführungen ist, dass als nächster Schritt der systematischen Erforschung des Distanzeffektes bei Ratschlagsgewichtung und Urteilssicherheit seine Überprüfung mit echten Ratschlägen, das heißt tatsächlich abgegebenen Schätzungen von anderen Individuen, erfolgen muss.

Drittens diente die Social Judgment Theorie (Sherif & Hovland, 1961) zwar als Argumentationsgrundlage für die Herleitung der Annahmen des Zusammenhangs zwischen Ratschlagsentfernung und –gewichtung; dennoch wurde nicht überprüft, ob individuelle *Latitudes of Acceptance* oder *Rejection* die Ursache der hier berichteten Befunde darstellen. Im Rahmen der Attitude Change-Forschung wurden bereits Ergebnisse dargelegt, die mit der SJT und ihren Annahmen teilweise nicht in Einklang gebracht werden konnten (Eagly & Telaak, 1972). Um zu überprüfen, in wieweit vorhandene Annahme- und Ablehnungsbereiche die Ratschlagsgewichtung tatsächlich beeinflussen, sollten daher weiterhin Untersuchungen durchgeführt werden, in denen für jedes zu treffende Urteil bzw. jeden Versuchsdurchgang auch Grenzwerte der Akzep-

tanz oder aber Plausibilität von Ratschlägen miterhoben werden. In diesem Zusammenhang ist auch nicht ganz auszuschließen, dass im JAS pro Aufgabe zwischen Plausibilitäts- und Akeptanzbereichen unterschieden werden kann. So ist von einem rationalen Standpunkt her davon auszugehen, dass bei den meisten quantitativen Schätzaufgaben Ratschläge hinsichtlich ihrer Plausibilität im Sinne von Richtigkeit bewertet und in Folge mehr oder weniger stark gewichtet werden. Im realen Leben zeigt sich jedoch oftmals auch, dass die Nutzung von Ratschlägen von ihrer Akzeptanz beim Rezipienten – im Sinne der Bereitschaft, sie anzunehmen – abhängt. Man kann hier beispielsweise an Situationen denken, in denen Ratschläge ungefragt erteilt werden – diese mögen dann vielleicht zwar inhaltlich plausibel sein und werden als solche auch vom Empfänger bewertet, werden aber unter Umständen trotzdem nicht angenommen bzw. gewichtet, weil es an der Bereitschaft dazu, sprich Akzeptanz, mangelt. Möglicherweise ist es daher sinnvoll, pro Aufgabe beide Arten von Bereichen zu erfragen.

Schließlich habe ich ähnlich wie die meisten Untersuchungen zum *Advice Taking* im Rahmen des JAS ein non-interaktives Setting zur Messung von Ratschlagsgewichtung und Sicherheitsveränderung genutzt. Obwohl dadurch eine hohe Vergleichbarkeit mit Referenzstudien erzielt werden konnte, limitiert diese Gegebenheit unter Umständen die Generalisierbarkeit der so ermittelten Ergebnisse auf reale Situationen, in denen Ratschläge erteilt und angenommen werden, da solche Situationen zumeist eine soziale Interaktion zwischen Judge und Advisor beinhalten. In diesen Situationen übt der Ratgeber dann nicht nur einen informationalen, sondern höchstwahrscheinlich auch einen normativen Einfluss auf die beratene Person aus (Deutsch & Gerard, 1955). Es wäre daher durchaus sinnvoll zu überprüfen, ob der Distanzeffekt auch bei direkter sozialer Einflussnahme durch den Advisor so konstant gezeigt wird wie in den hier berichteten Experimenten bzw. ob Probanden im Angesicht eines Ratgebers dessen insbesondere nahe Ratschläge anders gewichten (sie möglicherweise häufiger vollständig übernehmen, als sie eher zu Gunsten der eigenen ähnlichen Schätzung vollständig zu diskontieren) als wenn dieselben Personen ein verdecktes Finalurteil abgeben sollen.

8.6 Schlussbemerkung

Die im Rahmen meiner Dissertation berichteten Befunde zum Distanzeffekt bei Ratschlagsgewichtung und Urteilssicherheit erwiesen sich als systematisch und replizierbar und erweitern so einerseits den Begriff der Ratschlagsnutzung und andererseits das bisher bestehende Bild zur Annahme hinsichtlich des Zusammenhangs zwischen Ratschlagsdistanz und –gewichtung. Trotzdem können sie nur als eine erste Grundlage für das Verständnis dieser Beziehung angesehen werden. Viele Fragen und daraus entstehende Ansätze für zukünftige Untersuchungen sind noch unbeantwortet. Ist die distanzabhängige Gewichtung von Ratschlägen intendiert, und wenn ja, ist sie vielleicht sogar rational? Oder beeinträchtigt die distanzabhängige Gewichtung unter realen Ratschlagsbedingungen eher die Qualität getroffener Urteile und Entscheidungen? Wie könnte eine distanzabhängige Gewichtung unterbunden werden? Was ist ihre Ursache? Wo genau verläuft die Grenze zwischen einem als ähnlich und einem als identisch wahrgenommenen Ratschlag?

Mit Hinblick auf den praktischen Nutzen ist hierbei von großem Interesse, ob Distanzeffekte sich letztendlich positiv oder negativ auf die Urteilsqualität auswirken. Die von den Probanden vorwiegend angewandten Strategien der Wahl eines Urteils oder Bildung eines Mittelwertes aus den vorhandenen Optionen müssten demnach auch in Abhängigkeit der Ratschlagsdistanz in sinnvoller Häufigkeit jeweils an der richtigen Stelle eingesetzt werden. Wenn es gelänge, Menschen für den Zusammenhang zwischen Ratschlagsdiskrepanz und -qualität zu sensibilisieren, läge darin ein guter Ansatz für Interventionen mit dem Ziel der Vermittlung effektiver Strategien zur Ratschlagsnutzung.

9. Literatur

Aronson, E., Turner, J. A., & Carlsmith, J. M. (1963). Communicator credibility and communication discrepancy as determinants of opinion *change. Journal of Abnormal and Social Psychology, 67*, 31-36.

Bochner, S., & Insko, C. A. (1966). Communicator discrepancy, source credibility, and opinion change. *Journal of Personality and Social Psychology, 4,* 614-621.

Bonaccio, S., & Dalal, R. S. (2006). Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational Behavior and Human Decision Processes, 101,* 127-151.

Brewer, M. B., & Crano, W. D. (1968). Attitude change as a Function of Discrepancy and Source of Influence. *Journal of Social Psychology, 76*, 13-18.

Budescu, D. V. (2006). *Confidence in Aggregation of Opinions from Multiple Sources.* In Fiedler, K., & Juslin, P. (Eds.), Information Sampling and Adaptive Cognition (pp. 327-352). New York: Cambridge University Press.

Budescu, D.V., & Rantilla, A.K. (2000). Confidence in aggregation of expert opinions. *Acta Psychologica, 104*, 371-398.

Budescu, D. V., Rantilla, A. K., Yu, H., & Karelitz, T. K. (2003). The effects of asymmetry among advisors on the aggregation of their opinions. *Organizational Behavior and Human Decision Processes, 90,* 178–194.

Clement, R. W. & Krueger, J. (2000). The primacy of self-referent information in perceptions of social consensus. *British Journal of Social Psychology, 39,* 279 – 299.

Cohen, A. R. (1959) Communication discrepancy and attitude change. *Journal of Personality, 27,* 386-396.

Cross, R., Borgatti, S. P., & Parker, A. (2001). Beyond answers: dimensions of the advice network. *Social Networks, 23,* 215–235.

Cross, R., & Sproull, L. (2004). More Than an Answer: Information Relationships for Actionable Knowledge. *Organization Science, 14* (4), 446-462.

Deutsch, M., & Gerard, H. B. (1955). A Study of Normative and Informational Social Influences upon Individual Judgement. *The Journal of Abnormal and Social Psychology*, 51, 3, 629 – 636.

Eagly, A. H., & Telaak, K. (1972). Width of the latitude of acceptance as a determinant of attitude change. *Journal of Personality and Social Psychology, 23* (3), 388-397.

Feng, B., & MacGeorge, E. L. (2006). Predicting receptiveness to advice: Characteristics of the problem, the advice-giver, and the recipient. *Southern Communication Journal, 71,* 67-85.

Fisher, S., & Lubin, A. (1958). Distance as a determinant of influence in a two-person serial interaction situation. *Journal of Abnormal and Social Psychology, 56* (2), 230-238.

Freedman, J. L. (1964). Involvement, discrepancy, and change. *Journal of Abnormal and Social Psychology, 69* (3), 290-295.

Gardner, P.H., & Berry, D.C. (1995). The effect of different forms of advice on the control of a simulated complex system. *Applied Cognitive Psychology, 9,* 55-79.

Gigone, D., & Hastie, R. (1997). The impact of information on small group choice. *Journal of Personality and Social Psychology, 72*, 132–140.

Gino, F. (2008). Do we listen to advice just because we paid for it? The impact of advice cost on its use. *Organizational Behavior and Human Decision Processes, 107,* 234-245.

Gino, F., & Moore, D.A. (2007) Effects of Task Difficulty on Use of Advice. *Journal of Behavioral Decision Making, 20,* 21-35.

Gino, F., Shang, J., & Croson, R. (2009). The impact of information from similar or different advisors on judgment. *Organizational Behavior and Human Decision Processes, 108,* 287-302.

Harries, C., Yaniv, I., & Harvey, N. (2004). Combining advice: the weight of a dissenting opinion in the consensus. *Journal of Behavioral Decision Making, 17*, 333–348.

Harvey, N., & Fischer, I. (1997). Taking advice: Accepting help, improving judgment, and sharing responsibility. *Organizational Behavior and Human Decision Processes, 70*, 117-133.

Heath, C., & Gonzalez, R. (1995). Interaction with others increases decision confidence but not decision quality: evidence against information collection views of interactive decision-making. *Organizational Behavior and Human Decision Processes, 61,* 305–326.

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief adjustment model. *Cognitive Psychology,* 24, 1-55.

Hovland, C. I., Harvey, O. J., & Sherif, M. (1957). Assimilation and contrast effects in reactions to communication and attitude change. *The Journal of Abnormal and Social Psychology, 55* (2), 244-252.

Insko, C. A., Murashima, F., & Saiyadain, M. (1966). Communicator discrepancy, stimulus ambiguity, and influence. *Journal of Personality, 34*, 262-274.

Jungermann, H., & Fischer, K. (2005). *Using expertise and experience for giving and taking advice.* In T. Betsch & S. Haberstroh (Eds.),The routines of decision making (pp. 157‑173). Mahwah, NJ:Lawrence Erlbaum.

Krueger, J. I. (2003). Return of the ego—self-referent information as a filter for social prediction: Comment on Karniol (2003). *Psychological Review, 110*, 585–590.

Lim, J. S., & O'Connor, M. (1995). Judgemental Adjustment of initial forecasts: Its effectiveness and Biases. *Journal of Behavioral Decision Making, 8*, 149-168.

Lorenz, J., Rauhut, H., Schweitzer, F., & Helbing, D. (2011). How social influence can undermine the wisdom of the crowd effect. *Proceedings of the National Academy of Sciences, 108* (22), 9020-9025.

Minson, J. A., Liberman, V., & Ross, L. (2011). Two to Tango: Effects of Collaboration and Disagreement on Dyadic Judgment. *Personality and Social Psychology Bulletin, 37* (10), 1325-1338.

Mussweiler, T., & Strack, F. (2001). The semantics of anchoring. *Organizational Behavior and Human Decision Processes, 86*, 234-255.

Nemeth, C., & Endicott, J. (1976). The midpoint as an anchor: Another look at the discrepancy of position and attitude change. *Sociometry, 39* (1), 11-18.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science, 185*, 1124-1131.

Schrah, G.E., Dalal, R. S., & Sniezek, J. A. (2006).The adaptive decision maker is not an island: integrative expert advice with information search. *Journal of Behavioral Decision-Making, 19*, 43–60.

Schultze, T. (2010). *Die Übernutzung nicht valider Ratschläge: Warum schlechte Ratschläge über Gebühr berücksichtigt werden.* Dissertation, Biologische Fakultät, Georg-August-Universität Göttingen.

Schultze, T., Mojzisch, A., & Schulz-Hardt, S. (2012). On the inability to ignore useless advice: A case for anchoring in the Judge-Advisor-System. Unpublished manuscript.

Schulz-Hardt, S., Frey, D., Lüthgens, C., & Moscovici, S. (2000). Biased information search in group decision making. *Journal of Personality and Social Psychology, 78,* 655–669.

Schulz-Hardt, S., Jochims, M., & Frey, D. (2002). Productive conflict in group decision making: genuine and contrived dissent as strategies to counteract biased in-

formation seeking. *Organizational Behavior and Human Decision Processes, 88*, 563-586.

Sherif, M., & Hovland, C. (1961). *Social judgment: Assimilation and contrast effects in communication and attitude change.* New Haven: Yale University Press.

Sherif, M., Taub, D., & Hovland, C.I. (1958). Assimilation and contrasts effects of anchoring stimuli on judgments. *Journal of Experimental Psychology, 55* (2), 150-155.

Sniezek, J. A. (1992). Groups under Uncertainty: An Examination of Confidence in Group Decision Making. *Organizational Behavior and Human Decision Processes, 52,* 124-155.

Sniezek, J.A. (1999). *Judge Advisor Systems Theory and Research and Applications to Collaborative Systems and Technology.* Proceedings of the 32nd Hawaii International Conference on System Sciences (HICSS-32), Institute of Electrical and Electronics Engineers, Inc. (IEEE), Maui, Hawaii.

Sniezek, J. A., & Buckley, T. (1995). Cueing and cognitive conflict in judge-advisor decision making. *Organizational Behavior and Human Decision Processes, 62*, 159-174.

Sniezek, J. A., & Henry, R. A. (1989). Accuracy and confidence in group judgment. *Organizational Behavior & Human Decision Processes, 43*, 1–28.

Sniezek, J.A., & Henry, R.A. (1990). Revision, weighting and commitment in consensus group judgment. *Organizational Behavior and Human Decision Processes, 45,* 66–84.

Sniezek, J.A., Schrah, G.E., & Dalal, R.S. (2004). Improving judgment with prepaid expert advice. *Journal of Behavioral Decision Making, 17,* 173-190.

Sniezek, J. A., & Van Swol, L. M. (2001). Trust, confidence, and expertise in a Judge-Advisor System. *Organizational Behavior and Human Decision Processes, 84*(2), 288-307.

Soll, J.B., & Larrick, R.P. (2009). Strategies for revising judgment: How (and how well) people use others' opinions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35* (3), 780-805.

Soll, J.B., & Mannes, A.E. (2011). Judgmental aggregation strategies depend on whether the self is involved. *International Journal of Forecasting, 27*, 81 – 102.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124–1131.

Van Swol, L. M., & Sniezek, J. A. (2005). Factors affecting the acceptance of expert advice. *British Journal of Social Psychology, 44,* 443–461.

Wegener, D. T., Petty, R. E., Detweiler-Bedell, B. T., Blair, W., & Jarvis, 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.

Yaniv, I. (2004a). The Benefit of additional opinions. *Current Directions in Psychological Science, 13,* 75-78.

Yaniv, I. (2004b). Receiving other people's advice: Influence and benefit. *Organizational Behavior and Human Decision Processes, 93,* 1-13.

Yaniv, I., Choshen-Hillel, S. & Milyavsky, M. (2011). Receiving advice on matters of taste: Similarity, majority influence, and taste discrimination. *Organizational Behavior and Human Decision Processes, 115*, 111-120.

Yaniv, I., & Kleinberger, E. (2000). Advice taking in decision making: Egocentric discounting and reputation formation. *Organizational Behavior and Human Decision Processes, 83,* 260-281.

Yaniv, I., & Milyavsky, M. (2007). Using advice from multiple sources to revise and improve judgments. *Organizational Behavior and Human Decision Processes, 103,* 104-12.

Anhang

Anhang A: digitaler Anhang

Anhang B: Lebenslauf

Digitaler Anhang

Inhalt der Daten-CD:

- Dissertationsschrift als Datei (.docx und .pdf)
- SPSS-Daten
- Presentation®-Programmcodes für alle vier Experimente

Lebenslauf

Dipl.-Psych. Anne-Fernandine Rakotoarisoa

Universität Kassel

FB 01, Institut für Psychologie

Holländische Straße 36-38

34127 Kassel

Persönliches

Geburtstag: 27.10.1983

Geburtsort: Greifswald

Staatsangehörigkeit: Deutsch

Schulischer und wissenschaftlicher Bildungsgang

06/2002	Abitur (Gymnasium am Schloßteich, Chemnitz)
2002 – 2008	Studium der Psychologie (Diplom) an der Georg-August-Universität Göttingen
seit 2008	Promotion im Fach Psychologie an der Georg-August-Universität Göttingen