



**Allgemeine Psychologie
und Kognitionsforschung**
Justus-Liebig-Universität Gießen



Kollaboratives räumliches Schlussfolgern und Überzeugungsrevision

[Collaborative spatial reasoning and belief revision]

Inauguraldissertation
zur Erlangung des akademischen Grades
Doctor philosophiae (Dr. phil.)

an der

Justus-Liebig-Universität-Gießen
Fachbereich 06: Psychologie und Sportwissenschaft
Otto-Behaghel-Straße 10F
35394 Gießen

vorgelegt von

Andreas Reis, M.Sc. Psychologie

April, 2020

Betreuer und Erstgutachter: Prof. Dr. Markus Knauff

Zweitgutachter: PD. Dr. Kai Hamburger

Diese Arbeit wurde mit dem Projekt KN 465/6-2 der Deutschen Forschungsgemeinschaft (DFG), von Markus Knauff und Bernhard Nebel, innerhalb des Schwerpunktprogramms „New Frameworks of Rationality (SPP1516)“ gefördert.

Zusammenfassung

Es liegt in der Natur des Menschen, die Welt um sich herum zu entdecken und zu verstehen. Dazu ziehen wir fortwährend Schlussfolgerungen aus gegebenen Tatsachen und erlangen somit stetig neue Erkenntnisse. Das Schlussfolgern aus räumlichen Inhalten ist dabei eine spezielle Form des Denkens und hat sich in den letzten Jahrzehnten als eigenes Forschungsgebiet etabliert. Obwohl Kognitionswissenschaftler und Kognitionswissenschaftlerinnen sehr viel dazu beigetragen haben, die Prozesse beim räumlichen Schlussfolgern aufzuklären, haben sie einen wichtigen Einflussfaktor menschlichen Denkens vernachlässigt: die Gruppe. Das Ziel der vorliegenden Arbeit war deshalb zu untersuchen, wie mehrere Personen gemeinsam in einer Gruppe Schlussfolgerungen ziehen. In drei experimentellen Teilen habe ich jeweils eine von drei Phasen des räumlichen Schlussfolgerns (Konstruktion, Variation und Revision räumlicher mentaler Modelle) bei Gruppen getestet und mit den Leistungen von Individuen verglichen. Ich konnte zeigen, dass Gruppen analog zum individuellen Denken mentale Repräsentationen konstruieren und die Gruppenmitglieder dabei so viele Informationen wie möglich untereinander austauschen (Experiment 1 und 2). Bei der wichtigsten Phase des Schlussfolgerns – der Variation mentaler Modelle – waren Gruppen den Individuen überlegen (Experiment 3, 4 und 5). Selbst Gruppen, deren Mitglieder zuvor alleine nicht die Lösung finden konnten, waren durch Kollaboration in der Lage, eine valide Konklusion zu ziehen (Experiment 5). Zusätzlich zu den Befunden aus Gruppentestungen habe ich gezeigt, dass selbst ohne eine direkte Interaktion mit anderen Personen beim Revidieren räumlicher Überzeugungen ein Konformitätsdruck entstehen kann (Experiment 6) und die Vertrauenswürdigkeit der Quelle die Entscheidungen bei der Überzeugungsrevision erleichtern kann (Experiment 7). Meine Befunde zeigen also, dass der soziale Kontext ein wichtiger Aspekt für das menschliche Denken ist. Deshalb fasse ich meine Befunde in einer Theorie kollaborativen räumlichen Schlussfolgerns zusammen und zeige nicht nur, weshalb die Gruppe den Individuen beim Denken überlegen ist, sondern auch, wie Fehler beim kollaborativen Denken entstehen können.

Abstract

Trying to discover and to understand the world around us is deeply built into the nature of humankind. We constantly draw conclusions from given facts and thus gain new insights. A special form of this thinking is reasoning about spatial relations as part of spatial reasoning. Spatial reasoning has established as a separate field of research in recent decades. Although cognitive scientists have contributed a lot to explain the processes underlying spatial reasoning, they have neglected an important factor of human thought: the group. The aim of the present work was to examine how people draw conclusions together in a group. Therefore, I have tested the three phases of spatial reasoning (construction, variation and revision) of spatial mental models in groups and compared their performances with individuals. On the one hand, I show that groups construct mental representations analogously to individual thinking. On the other hand, group members try to exchange as much information as possible with each other (experiments 1 and 2). In the crucial phase of reasoning – the variation of mental models – groups outperformed individuals in reasoning (experiment 3, 4 and 5). Even groups whose members could not find a solution on the individual level were able to draw a valid conclusion as a group (Experiment 5). In addition to the findings from group testing, I have shown that – even without direct interaction with other persons – a pressure of conformity can arise when revising spatial beliefs (Experiment 6) and the source trustworthiness may facilitate revision decisions (Experiment 7). My findings show that the social context is an important aspect of human thinking. Thus, at the end, I integrate my findings in a theory of collaborative spatial reasoning. With this dissertation, I not only demonstrate why the group is superior to individuals in thinking but also show how errors in collaborative thinking can occur.

Inhaltsverzeichnis

Einleitung	1
Kapitel 1: Theoretischer Hintergrund	5
1.1 Logisches Denken	5
1.2 Räumliches Schlussfolgern	7
1.2.1 Theorien zum räumlichen Schlussfolgern	8
1.2.2 Psychologische Befunde zum räumlichen Schlussfolgern	15
1.3 Überzeugungsrevision	21
1.3.1 Psychologische Befunde zur Überzeugungsrevision	23
1.3.2 Räumliche Überzeugungsrevision	28
1.4 Gruppenprozesse	32
1.4.1 Entscheidungsfindung und Urteilen in Gruppen	33
1.4.2 Problemlösen in Gruppen	35
1.4.3 Psychologische Befunde zu Gruppenprozessen	44
Kapitel 2: Integration der Theorien und Hypothesen	53
2.1 Kollaborative Konstruktion mentaler Modelle	56
2.2 Kollaborative Variation mentaler Modelle	58
2.3 Soziale Revision mentaler Modelle	61
2.4 Hypothesen	62
Kapitel 3: Experimenteller Teil 1 - Kollaborative Konstruktion mentaler Modelle	64
3.1 Experiment 1 – Kollaborative Konstruktion mentaler Modelle	65
3.1.1 Methoden	65
3.1.2 Ergebnisse	68
3.1.3 Diskussion	71
3.2 Experiment 2 – Kollaborative Konstruktion und Teilen mentaler Modelle	73
3.2.1 Methoden	73

3.2.2	Ergebnisse	76
3.2.3	Diskussion	77
3.3	Zusammenfassung Kapitel 3	79
Kapitel 4: Experimenteller Teil 2 - Kollaborative Variation mentaler Modelle		81
4.1	Experiment 3: Kollaborative Variation mit realen Objekten	82
4.1.1	Methoden.....	82
4.1.2	Ergebnisse	87
4.1.3	Diskussion	90
4.2	Experiment 4: Kollaborative Variation ohne reale Objekte	92
4.2.1	Methoden.....	92
4.2.2	Ergebnisse	94
4.2.3	Diskussion	97
4.3	Experiment 5: Kollaborative Variation und Reihenfolgeeffekte.....	99
4.3.1	Methoden.....	99
4.3.2	Ergebnisse	102
4.3.3	Diskussion	106
4.4	Zusammenfassung Kapitel 4	108
Kapitel 5: Experimenteller Teil 3 - Soziale Revision mentaler Modelle....		112
5.1	Experiment 6: Konformität und räumliche Überzeugungsrevision.....	113
5.1.1	Methoden.....	114
5.1.2	Ergebnisse	116
5.1.3	Diskussion	118
5.2	Experiment 7: Quellenprinzip und räumliche Überzeugungsrevision	119
5.2.1	Vorstudie	120
5.2.2	Methoden.....	121
5.2.3	Ergebnisse	124
5.2.4	Diskussion	126

5.3 Zusammenfassung Kapitel 5	129
Kapitel 6: Allgemeine Diskussion	131
6.1 Kollaboratives räumliches Schlussfolgern - Eine Theorie des Denkens in Gruppen.	132
6.1.1 Aufbau der KRS Theorie.....	133
6.1.2 Die Gruppe als epistemischer Wetzstein.....	136
6.1.3 Fehler beim kollaborativen räumlichen Schlussfolgern	140
6.2 Konsequenzen für die Denkpsychologie	141
6.2.1 Andere Denkprobleme	142
6.2.2 Alternative Theorien zum Denken	143
6.3 Konsequenzen für die Sozialpsychologie.....	148
6.4 Konsequenzen für die Gesellschaft	151
6.5 Weiterführende Forschung	153
6.5.1 Aufgabenschwierigkeit und weitere Effekte beim Denken.....	153
6.5.2 Überzeugungsrevision in Gruppen	155
6.5.3 Gruppengröße und Reihenfolgeeffekte	156
6.5.4 Soziale Entscheidungsschemata	158
6.5.5 Synergieeffekte beim kollaborativen Denken	159
6.6 Fazit.....	161
Kapitel 7: Literatur	162

Einleitung

Sherlock Holmes ist bekannt für seinen scharfen Verstand. Er überrascht sein Umfeld jedes Mal aufs Neue mit scheinbar unmöglichen Schlussfolgerungen. Man könnte meinen, dass Holmes für die Lösung seiner Fälle lediglich seinen eigenen Verstand benötigt. Doch welche Rolle spielt dann sein ständiger Begleiter, Dr. John Watson? In dem Roman „*The Adventure of the Creeping Man*“ versucht Watson selbst eine Antwort auf diese Frage zu geben:

„Ich war ein Wetzstein für seinen Geist. Ich stimulierte ihn. Er dachte gerne laut in meiner Gegenwart. Zwar konnte man seine Bemerkungen kaum als an mich gerichtet bezeichnen, – viele von ihnen hätten ebenso gut seiner Bettstatt gelten können – aber da es nun mal zur Gewohnheit geworden war, hatte es sich in mancher Hinsicht als hilfreich erwiesen, dass ich alles registrierte und mit Einwürfen bedachte. Wenn ihn eine gewisse Langsamkeit meiner Denkvorgänge auch irritierte, so hatte dies zur Folge, dass seine Eingebungen und Ideen umso lebhafter und rascher aufflammten und blitzten. So also sah meine bescheidene Rolle in unserem Bündnis aus.“

– Arthur Conan Doyle, Sherlock Holmes: Adventure of the Creeping Man (aus: The Casebook of Sherlock Holmes, 1993)

Watson hat gelernt, Holmes mit den richtigen Fragen auf die richtigen Ideen zu bringen. Kann es also sein, dass nicht ausschließlich der Verstand von Sherlock Holmes, sondern vielmehr die Interaktion zwischen den beiden Detektiven die herausragenden Denkfähigkeiten erklärt? Diese Annahme scheint nicht abwegig zu sein, wenn man bedenkt, dass jeder Mensch fortwährend mit anderen Personen interagiert. Sei es ein banales Gespräch zwischen zwei Kollegen¹ am Kaffeeautomaten oder eine Gruppe von Wissenschaftlern, die eine Mondlandung koordinieren. Wir begeben uns ständig in Gruppen und versuchen gemeinsam die Welt zu verstehen und zu verändern. Dabei ist es für uns selbstverständlich, dass die Gruppe rationalere Entscheidungen als der Einzelne trifft. Deshalb überlassen wir kleinen ausgewählten Gruppen die Verantwortung, Antworten für die komplexesten Probleme wie die Klimakrise zu finden. Für den Anthropologen und Psychologen Robin Dunbar ist die Gruppe so bedeutend, dass er sie als einen der wichtigsten Faktoren der menschlichen Phylogenese ansieht. So konnte der Mensch nach Dunbar erst durch die Interaktion in Gruppen ein Gehirn entwickeln, welches zu höheren Denkleistungen in der Lage ist (Dunbar, 1998). Das Denken in Gruppen scheint also einen sehr hohen Stellenwert für den Menschen

¹ In dieser Arbeit wird aus Gründen der besseren Lesbarkeit ausschließlich die männliche Form verwendet. Sie bezieht sich bei allen Ausführungen auf Personen beiderlei Geschlechts.

zu haben. Doch wie denken Menschen in Gruppen? Und unterscheiden sich die Denkprozesse in Gruppen von denen einzelner Personen?

Um die Prozesse menschlichen Denkens zu untersuchen, bietet sich vor allem das *räumliche Schlussfolgern* an. Beim räumlichen Schlussfolgern nutzen Menschen ihr vorhandenes Wissen über die Welt, um Informationen abzuleiten. Wenn wir wissen, dass Nils links von Jan ist und Jan links von Sabrina, können wir – obwohl es nicht explizit geäußert wurde – daraus ableiten, dass Nils auch links von Sabrina sein muss. Dieses Denken gilt als Grundbaustein komplexer kognitiver Prozesse (Halford, Wilson, & Phillips, 2010) und ist so allgegenwärtig, dass wir oft gar nicht bemerken, wenn wir aus räumlichen Inhalten andere Informationen ableiten. So folgen wir im Straßenverkehr automatisch der Regel „rechts vor links“ und ordnen sogar ganze politische Lager räumlich in „rechts“ oder „links“ ein. Das räumliche Denken kann also auch weit über simple Richtungsanweisungen hinaus dazu dienen, Zusammenhänge zwischen Personen, Objekten oder Gruppen zu bilden und diese ausreichend von anderen abzugrenzen (Tversky, 2011). Obwohl das räumliche Schlussfolgern so allgegenwärtig ist und wir mit Leichtigkeit Schlüsse aus räumlichen Gegebenheiten ziehen können, hat es viele Jahrzehnte gedauert, die Prozesse, die diesem Denken unterliegen, zu erklären. Dabei konnte man zeigen, dass wir auch in experimentellen Tests sehr gut darin sind, Schlussfolgerungen mit räumlichen Inhalten durchzuführen. Es gibt aber auch Fälle, die so komplex sind, dass wir manchmal an unsere Grenzen stoßen. Solche schwierigen Aufgaben geben Kognitionswissenschaftlern die Möglichkeit, genaue Aussagen darüber zu machen, wie Fehler beim Denken entstehen und wir können sehr viel darüber lernen, wie diese menschlichen Schlussfolgerungen funktionieren. Interessanterweise basieren aber alle diese Befunde auf Experimenten mit einzelnen Personen. Wenn jedoch die Gruppe für das menschliche Denken so wichtig ist und der Mensch fortwährend Schlüsse aus räumlichen Informationen zieht, dann sollte das räumliche Schlussfolgern auch in Gruppen untersucht werden.

Das Ziel der vorliegenden Arbeit soll deshalb sein, Aufschlüsse darüber zu geben, wie Gruppen beim räumlichen Schlussfolgern denken. Dazu habe ich Gruppen mit räumlichen Denkaufgaben konfrontiert und sie diese kollaborativ – genau wie Sherlock Holmes und Dr. Watson – lösen lassen. Ich habe getestet, wie Gruppen gemeinsam eine Repräsentation von räumlichen Problemen bilden und wie sie diese weiterverarbeiten. Dabei fragte ich mich, ob es der Gruppe – analog zum individuellen Denken – auch Probleme bereitet, wenn sie viele Informationen auf einmal zum korrekten Schlussfolgern berücksichtigen muss. Es gibt bereits Theorien aus anderen Bereichen der Kognitionsforschung die annehmen, dass die Gruppe

durch die Kombination ihrer individuellen Überlegungen mehr Informationen als das Individuum verarbeiten kann. Doch trifft dies auch beim räumlichen Schlussfolgern zu? Wenn es der Gruppe gelingt, mehr Informationen zu berücksichtigen und in eine gemeinsame Antwort zu kombinieren, sollte die Gruppe auch beim räumlichen Schlussfolgern ein wichtiges epistemisches Werkzeug darstellen. Das bedeutet, dass die Gruppe – von Dr. Watson als „Wetzstein für den Geist“ bezeichnet – das folgerichtige Denken fördern kann. Ich werde also in meiner Arbeit nicht nur testen ob, sondern auch wie die Gruppe Informationen kombiniert und zu einer Schlussfolgerung gelangt.

Um meine Hypothesen zum kollaborativen Denken abzuleiten, ist es wichtig zu verstehen, wie das individuelle Denken funktioniert und wie diese Prozesse mit dem Denken in Gruppen zusammenhängen können. Dazu habe ich das *Kapitel 1: Theoretischer Hintergrund* in zwei Abschnitte unterteilt. Im ersten Abschnitt ordne ich das räumliche Schlussfolgern historisch in das logische Denken ein. Anschließend fasse ich die Theorien und Befunde zum individuellen räumlichen Schlussfolgern zusammen und zeige, welche Prozesse wir Menschen beim Denken verwenden. Im zweiten Abschnitt stelle ich Theorien und Befunde zu Gruppenprozessen dar. Da bisher kaum Experimente zum Schlussfolgern in Gruppen durchgeführt wurden, habe ich mich hierbei auf einen anderen, sehr gut untersuchten Bereich der Kognitionsforschung bezogen: das kollaborative Problemlösen. Das Problemlösen stellt ebenfalls eine höhere kognitive Funktion dar und kann uns theoretische, aber auch methodische Hinweise darüber geben, wie Gruppen gemeinsam zu einer Lösung gelangen. Nachdem ich also das individuelle räumliche Schlussfolgern und das Problemlösen in Gruppen im ersten Kapitel noch getrennt behandle, werde ich in *Kapitel 2: Integration der Theorien und Hypothesen* die Theorien aus beiden Bereichen zusammenführen und daraus meine Hypothesen ableiten. Diese Hypothesen habe ich in insgesamt drei experimentellen Teilen getestet.

In *Kapitel 3: Kollaborative Konstruktion mentaler Modelle* habe ich in zwei Experimenten untersucht, wie die Gruppe gemeinsam eine Repräsentation einer räumlichen Aufgabe (wie das Beispiel mit Nils, Jan und Sabrina) bildet und wie die Gruppenmitglieder die Informationen miteinander austauschen. Das *Kapitel 4: Kollaborative Variation mentaler Modelle* ist der Schwerpunkt meiner experimentellen Arbeit. In drei Experimenten habe ich sehr schwierige Aufgaben an Gruppen getestet. Bei diesen Aufgaben müssen sehr viele Informationen berücksichtigt werden. Dies führt dazu, dass die Gruppe mehr als eine mögliche Antwort berücksichtigen und eine ursprünglich konstruierte Repräsentation in Hinblick auf die weiteren Möglichkeiten verändern, also variieren muss. Im *Kapitel 5: Soziale*

Revision mentaler Modelle habe ich einen besonderen Fall beim räumlichen Schlussfolgern untersucht. Dabei habe ich in zwei Experimenten getestet, wie Menschen sich verhalten, wenn sie eine zunächst korrekte räumliche Repräsentation gebildet haben, diese aber aufgrund einer neuen Information (z. B. der Aussage eines Gruppenmitglieds) überdacht werden musste. Das bedeutet, dass eine Überzeugung revidiert werden muss. Hierbei habe ich untersucht, wie soziale Einflüsse diese *räumliche Überzeugungsrevision* verändern können.

Die Befunde meiner Experimente werde ich im *Kapitel 6: Allgemeine Diskussion* in eine Theorie zum *kollaborativen räumlichen Schlussfolgern* integrieren. Mit dieser Theorie zeige ich nicht nur, warum Dr. Watson eine sehr gute Idee über seine Rolle in dem literarischen Duo hat, sondern auch, warum sich bei den zwei Detektiven trotzdem manchmal doch ein paar Fehler beim Denken einschleichen können.

Kapitel 1: Theoretischer Hintergrund

Das Schlussfolgern mit räumlichen Relationen gehört zu den grundlegendsten Prozessen menschlichen Denkens. So ist es wichtig zum Bestimmen von Entitäten im Raum (Denis & Cocude, 1989), zum Planen von Routen (Johnson-Laird & Byrne, 1991) und sogar dann, wenn wir soziale Kategorien bilden (Tversky, 2011). Deshalb ist es nicht verwunderlich, dass das räumliche Denken eines der am besten untersuchten Domänen menschlicher Kognitionen darstellt. Doch wie schlussfolgern Gruppen mit räumlichen Inhalten? Um sich einer Antwort dieser Frage zu nähern, ist es notwendig, zunächst das räumliche Schlussfolgern beim Einzelnen zu betrachten. Dabei wurde das räumliche Schlussfolgern nicht isoliert von anderen Bereichen des Denkens untersucht. Deshalb ist es wichtig, das Denken mit räumlichen Inhalten in den größeren Kontext des *logischen Denkens* einzuordnen.

Im folgenden Abschnitt werde ich zunächst die philosophischen und psychologischen Ursprünge des logischen Denkens beschreiben. Dann werde ich auf die psychologischen Theorien zum räumlichen Schlussfolgern bei Individuen eingehen und schließlich die Theorien und Forschungen zu Gruppenprozessen darlegen. Da bisher kaum Untersuchungen zum Schlussfolgern in Gruppen durchgeführt wurden, werde ich vor allem Befunde aus dem Bereich des Problemlösens in Gruppen berichten. Dabei habe ich mich auf ein Problem fokussiert, für dessen Lösung bereits viele Anteile schlussfolgernden Denkens notwendig sind.

1.1 Logisches Denken

Unter logischem Denken versteht man das Ableiten von Informationen aus vorher gegebenen Aussagen. Es unterteilt sich in *Induktion* und *Deduktion*. Mit *induktivem Denken* ist das Schließen von einer speziellen Situation auf eine allgemeine Regel gemeint. Das bekannteste Beispiel für induktives Denken ist vermutlich die Wissenschaft selbst. Durch einzelne Beobachtungen werden Rückschlüsse auf allgemeine Prinzipien gezogen. Es ist jedoch nie gesichert, ob eine andere Beobachtung möglicherweise auf andere Ergebnisse und somit zu unterschiedlichen Schlussfolgerungen kommt. Würde man zum Beispiel den Straßenverkehr in Großbritannien beobachten, könnte man schließen, dass der Linksverkehr eine allgemeine Gesetzmäßigkeit im weltweiten Straßenverkehr darstellt. Wie aber jeder weiß, ist Großbritannien nicht die Regel, sondern zusammen mit wenigen anderen Ländern eine Ausnahme. Es wäre in diesem Beispiel also falsch, vom Einzelfall auf die Gesamtheit zu schließen. Deshalb sind induktive Schlüsse nie eindeutig sichere Schlüsse. Beim *deduktiven Denken* hingegen ist genau das Gegenteil der Fall. Hier werden Schlussfolgerungen aus

allgemeinen Aussagen auf spezifische Situationen durchgeführt. Die Deduktion ist ein Prozess bei dem keine neue Information generiert wird. Es wird lediglich abgeleitet, was implizit in den Aussagen enthalten ist und somit sind sichere Schlüsse möglich. Historisch geht die Deduktion auf die klassische Logik von Aristoteles zurück. Demnach deckt man allein mit Hilfe der Struktur von Argumenten Widersprüche und Zusammenhänge auf. Es wird immer von *Prämissen* (Aussagen) auf eine *Konklusion* (Schlussfolgerung) geschlussfolgert. Eine wichtige Annahme nach der klassischen Logik ist, dass eine Konklusion immer wahr ist, wenn die Prämissen wahr sind.

Das Ziel bei der Erforschung der Deduktion ist herauszufinden, wie – also mit welchen Prozessen – der Mensch logisch (richtig) denkt. Dazu verwenden Psychologen die nach der klassischen Logik abgeleiteten Schlussfolgerungen als normatives Bezugssystem (Chater & Oaksford, 2012). Das bedeutet, dass sie die Antworten der Menschen in Experimenten mit den formal korrekten Konklusionen nach der klassischen Logik vergleichen. Um normativ korrekte Schlussfolgerungen zu bestimmen, werden verschiedene Logiken verwendet. Diese werden in die *Aussagenlogik* und *Prädikatenlogik* unterteilt. Die Aussagenlogik beschäftigt sich mit so genannten *Konditionalen* (Wenn..., dann... Aussagen). Das heißt, dass sie Sätze mit *Implikationen* untersuchen. Die Prädikatenlogik befasst sich mit *Syllogismen* (Alle X sind Y) und *Relationen* (X ist größer als Z). Zu letzteren zählt auch das räumliche Schlussfolgern mit Aussagen wie „Das Fahrrad ist links von dem Haus“. Das Schlussfolgern mit Relationen ist ein Spezialfall der allgemeinen Prädikatenlogik und wird deshalb als *linearer Syllogismus* bezeichnet (Knauff, 2006). Das bedeutet also, dass das räumliche Schlussfolgern zur Deduktion gehört und ein epistemischer Prozess ist, bei dem von vorhandenem Wissen auf weitere Gegebenheiten geschlussfolgert wird.

Räumliches Schlussfolgern hat sich in den letzten Jahrzehnten als eigenständiges Forschungsgebiet mit eigenen Theorien und Methoden etabliert. Deshalb möchte ich im folgenden Abschnitt zeigen, wie das räumliche Schlussfolgern theoretisch in die Erforschung des Denkens einzuordnen ist und gebe anschließend einen Überblick über die aktuellen Befunde aus psychologischen Untersuchungen.

1.2 Räumliches Schlussfolgern

Eine Aufgabe zum räumlichen Schlussfolgern hat folgende Struktur:

(1) Der Apfel ist links von der Birne (Ar_1B)

(2) Die Birne ist links von der Mango. (Br_2C)

Konklusion: Ist der Apfel links von der Mango? (Ar_3C)

Die Aufgabe besteht aus zwei Prämissen und einer Konklusion. Eine Konklusion kann dabei wie im Beispiel als Frage formuliert sein oder muss als zusammenhängende Anordnung (Apfel Birne Mango) selbst angegeben werden. Eine Prämisse beim räumlichen Schlussfolgern besteht in der Regel aus zwei Objekten. In der Logik nennt man diese Objekte *Argumente* oder auch *Terme*. Außerdem enthält eine Prämisse eine Relation, die man in der Logik als *Prädikat* bezeichnet. Eine Aufgabe zum räumlichen Schlussfolgern wird nach der Anzahl der enthaltenen Terme benannt (n-Term Aufgabe). In diesem Beispiel wäre es also eine *drei-Term* Aufgabe. Prämissen werden formal mit Hilfe von Klein- und Großbuchstaben dargestellt (ArB). Die Großbuchstaben (A und B) stellen dabei die Terme dar. Die Kleinbuchstaben (r) stehen für das Prädikat. Bei mehreren Prädikaten wird das Zeichen mit einer Nummerierung versehen (z. B. Ar_1B , Br_2C). Prädikate haben eine sogenannte *Stelligkeit*. Ein Prädikat wie „links von“ bringt zwei Terme (ArB) in Zusammenhang. Deshalb wird es „zweistellig“ genannt. Prädikate können aber theoretisch beliebig viele Terme enthalten. Sie sind dabei jedoch immer auf eine endliche Menge beschränkt (Knauff, 2006). So kann das Prädikat „zwischen“ Aussagen über drei oder mehr Terme treffen (z. B. „A ist zwischen B und C“). In der Regel werden für die Untersuchung räumlichen Schlussfolgerns zweistellige Prädikate verwendet. Diese können nach Eigenschaften der *Reflexivität*, *Symmetrie* und *Transitivität* unterschieden werden (Goodwin & Johnson-Laird, 2005). Reflexivität bedeutet, dass ein Term Bezug zu sich selbst nimmt (ArA) und hat bei räumlichen Relationen kaum eine Relevanz. Symmetrische Prädikate wie „A ist neben B“ lassen den Schluss zu, dass „B neben A“ ist. Man kann die Terme im Satz also beliebig tauschen. Bei einem asymmetrischen Prädikat wie „rechts von“ ist dies nicht der Fall. Transitivität beschreibt die Eigenschaft, dass aus zwei Prädikaten wie „ Ar_1B “ und „ Br_1C “ automatisch das dritte Prädikat „ Ar_1C “ folgt. Man bezeichnet die Konklusion deshalb auch als *Komposition* der ersten beiden Prämissen (Knauff & Knoblich, 2017). William James (1890) behandelt das Konzept der Transitivität in seinem *Principle of Mediate Comparison*. Dieses in

der Logik als das *fundamentale Inferenzprinzip* bekannte Prinzip besagt, dass man bei einer linearen Anordnung mit beliebig vielen Termen ($A < B < C < \dots < Z$) beliebig viele Zwischenelemente entfernen kann, ohne dass sich die Relationen zu den verbleibenden Termen ändern. Schon damals sah James das Denken mit transitiven Relationen als das womöglich wichtigste Gesetz menschlichen Denkens (James, 1890). Es gibt aber auch Relationen, die keine transitiven Schlüsse zulassen. Ein Prädikat wie „ist Mutter von“ ist intransitiv, da man aus den zwei Aussagen „Julia ist Mutter von Anja“ und „Anja ist Mutter von Sven“ nicht automatisch „Julia ist Mutter von Sven“ schlussfolgern kann (Evans, Newstead, & Byrne, 1993).

Die Prädikatenlogik kann uns sagen, wie Aufgaben zum räumlichen Schlussfolgern formal dargestellt werden und wie verschiedene Prädikate nach den Eigenschaften der Reflexivität, Symmetrie und Transitivität unterschieden werden. Daraus können wir formal korrekte Schlussfolgerungen ableiten und somit vorgeben, welchen Normen der Mensch beim Denken folgen sollte. Diese Normen sagen aber noch nichts darüber aus, welchen Prozessen das Denken beim Schlussfolgern unterliegt. Deshalb werden im nächsten Abschnitt die psychologischen Theorien zum räumlichen Schlussfolgern dargestellt.

1.2.1 Theorien zum räumlichen Schlussfolgern

Obwohl das räumliche Schlussfolgern so wichtig für den Menschen ist, hat es viele Jahre gedauert, diese Art des Denkens zu erklären. Dabei haben sich in der Psychologie zahlreiche Theorien gebildet, die das räumliche Schlussfolgern beschreiben. Diese lassen sich allgemein unter den *Theorien Mentaler Logik (TML)* und *Theorien Mentaler Modelle (TMM)* zusammenfassen. Beide Theorien nehmen an, dass Menschen durch das Verarbeiten mentaler Repräsentationen schlussfolgern (Knauff, 2006). Sie unterscheiden sich jedoch grundsätzlich in ihren Annahmen über die Art der Repräsentation und den Prozessen der Verarbeitung. Das bedeutet, dass sie nicht nur unterschiedliche Aussagen darüber treffen, wie das Denken verläuft, sondern auch, wie Fehler entstehen. Inzwischen sind sich die Forscher überwiegend einig, dass das räumliche Schlussfolgern am besten durch die Modelltheorien beschrieben wird (Goodwin & Johnson-Laird, 2005; Knauff, 2006, 2013; Knauff & Knoblich, 2017; Vandierendonck, Dierckx, & De Vooght, 2004). Trotzdem hatte die TML einen starken Einfluss auf experimentelle Untersuchung räumlichen Schlussfolgerns. So wurden zum Beispiel ganze Aufgaben zum räumlichen Schlussfolgern so konstruiert, um die Vorhersagen der TML mit der TMM zu vergleichen. Deshalb ist es auch für das Erforschen von Gruppenleistungen wichtig, beide Theorien zu verstehen und historisch einzuordnen.

1.2.1.1 Theorie mentaler Logik

Die Basis der TML findet sich bereits in den Arbeiten von Piaget (1965). Dieser nahm an, dass das Schlussfolgern beim Menschen anhand von Regeln abläuft. Diese Regeln müssen sich bis zur frühen Adoleszenz (Stufe formal operationalen Denkens, ab 11 Jahre) entwickeln und können dann für ein beliebiges Problem auf abstrakte Repräsentationen angewendet werden. Diese Idee wurde von Rips (1994) und Brain und O'Brien (1991) aufgegriffen und weiterentwickelt. Nach ihren Theorien ist eine Konklusion nur dann korrekt, wenn sie eindeutig aus der Struktur der Prämissen ableitbar ist (Gentzen, 1935) und entspricht somit einer logisch-mathematischen Sicht auf das Denken. Deshalb spricht man nach der TML auch von der *Ableitbarkeit* einer Konklusion (Rips, 1994). Das Denken läuft nach der TML in drei Schritten ab. Zunächst wird die syntaktische Struktur der Prämissen im Arbeitsgedächtnis encodiert. Das heißt, dass die ursprünglichen Sätze in Form von sprachnahen Propositionen repräsentiert werden. Eine Proposition ist eine kleinstmögliche Wissenseinheit, deren Inhalt entweder wahr oder falsch sein kann (Knauff, 2006). Propositionen werden amodal – also unabhängig von der Modalität (visuell, auditiv, haptisch, emotional) – repräsentiert (Kaup & Dudschig, 2017). Eine sprachnahe propositionale Repräsentation bedeutet also nach der TML, dass die Prämissen als kleinstmögliche Einheiten repräsentiert werden und dabei die syntaktische Struktur der Prämissen erhalten bleibt. Die *Semantik* (Bedeutung) der Prämissen spielt nach der TML keine Rolle. Bei der Prämisse „der Apfel ist links von der Birne“ ist also nur die logische Form [links(Apfel,Birne)] relevant. Deshalb spricht man bei der TML auch von den *syntaktischen Theorien*. Auf die Repräsentationen werden im zweiten Schritt Regeln angewandt. Hierbei werden die Prämissen – ähnlich wie beim mathematischen Ableiten – immer weiter umgeformt und reduziert. Im dritten Schritt wird die abgeleitete Information wieder in ein sprachliches Format zurückgeführt. Nach der TML können Fehler beim logischen Denken auf drei Arten zurückzuführen sein (Rips, 1994). *Verständnisfehler* entstehen, wenn die Prämissen fehlerhaft ins Arbeitsgedächtnis encodiert werden. Ein Grund für Verständnisfehler kann durch die Grice'sche Implikationshypothese (Grice, 1975) veranschaulicht werden. Diese Hypothese besagt, dass Personen Aussagen wie „A ist links von B“ als „A ist *direkt links* von B“ verstehen. Sie implizieren also, dass eine Aussage mehr beinhaltet als ihre rein logische Bedeutung. *Prozessfehler* beschreiben den Verlust von Informationen aus dem Arbeitsgedächtnis. Diese werden vor allem durch Ablenkung oder Konzentrationsmangel erklärt. Bei *Strategiefehlern* gelingt es den Personen nicht, die unterschiedlichen Regeln, die für eine Schlussfolgerung notwendig sind, adäquat auszuwählen

und auf die Repräsentationen anzuwenden. Allgemein lässt sich also sagen, dass mehr Fehler entstehen, wenn mehr Schritte zum Ableiten der Informationen notwendig sind.

Die TML von Rips orientiert sich sehr stark an der formalen Logik und wurde überwiegend für das konditionale Schlussfolgern getestet. Um aber Inferenzen beim relationalen Schlussfolgern zu beschreiben, entwickelte Hunter (1957) ein spezielles, an dem *fundamentalen Inferenzprinzip* orientiertes System. Ziel ist, die Prämissen in ein einheitliches System zu überführen. Dieses Ziel ist erreicht, wenn alle Prämissen einer einheitlichen Richtung (z. B. links von) folgen und aufeinander aufbauen. Dadurch werden transitive Inferenzen möglich.

Beispiel

- (1) Die Birne ist links von der Mango. (Br₂C)
- (2) Die Birne ist rechts von dem Apfel. (Br₁A)

Um für dieses Beispiel ein einheitliches Schema zu bilden, sind laut Hunter zwei Operationen (Regeln) notwendig. Die erste ist das *Umkehren* von Prämissen. Das heißt, dass man aus „Die Birne ist rechts von dem Apfel.“ durch Umkehr „Der Apfel ist links von der Birne“ bilden kann. Die zweite notwendige Operation ist der *Tausch* der Prämissen. Es macht nach Hunter bei drei-Term Aufgaben keinen Unterschied, welche Prämisse zuerst präsentiert wird und man kann durch einfaches Vertauschen der Prämissen ein einheitliches Schema bilden. Nun ist ohne weitere Schritte die transitive Inferenz „Der Apfel ist links von der Birne“ möglich. Hunters Theorie kann Inferenzen für eindimensionale drei-Term Aufgaben sehr gut beschreiben. Sie bieten jedoch kein adäquates System für zweidimensionale Aufgaben wie das folgende Problem:

- (1) Der Apfel ist links von der Birne (Ar₁B)
- (2) Die Mango ist unter der Birne (Cr₂B)

Konklusion: Ist der Apfel links von der Mango? (Ar₃C)

Ein formales System mit Regeln für mehrdimensionales räumliches Schlussfolgern wurde von Hagert und Hansson (1983) und Hagert (1985) vorgeschlagen. Dieses System beginnt ebenfalls damit, dass die Prämissen in sprachnaher propositionaler Form im Arbeitsgedächtnis encodiert und dort durch Regeln manipuliert werden. Er erweitert jedoch

das System von Hunter um zusätzliche Regeln. Die wichtigsten sind dabei die der *Konjunktion* und *Disjunktion*. Unter Konjunktion versteht man das Zusammenführen von zwei oder mehr Prämissen. So kann man aus „Der Apfel ist links von der Birne“ und „Die Mango ist unter der Birne“ den Satz „Die Mango ist unter der Birne, und der Apfel ist links von ihnen“ bilden. Syntaktisch wird dies folgendermaßen dargestellt:

$$\text{links}(\text{Apfel}, \text{Birne}) \wedge \text{unter}(\text{Mango}, \text{Birne}) \rightarrow \text{links}(\text{Apfel}, \text{unter}(\text{Mango}, \text{Birne}))$$

Dies lässt die einfachere Inferenz ($\text{links}(\text{Apfel}, \text{Mango})$) zu. Durch Rückführen des syntaktischen in ein sprachliches Format wird dann „Der Apfel ist links von der Mango“. Die wichtigsten Regeln nach Hagert sind in Tabelle 1 aufgelistet.

Tabelle 1

Syntaktische Regeln zum räumlichen Schlussfolgern nach Hagert (1985)

Regel	Formale Darstellung
Umkehr	$\text{links}(x,y) \leftrightarrow \text{rechts}(y,x)$
Negation	$\text{links}(x,y) \rightarrow \neg \text{rechts}(x,y)$
Konjunktion	$\text{links}(x,y) \wedge \text{unter}(z,y) \rightarrow \text{links}(x, \text{unter}(z,y))$
Disjunktion	$\text{links}(x, \text{unter}(z,y)) \rightarrow \text{links}(x,y) \wedge \text{unter}(z,y)$
Transitivität	$\text{links}(x,y) \wedge \text{links}(y,z) \rightarrow \text{links}(x,z)$

Anmerkung: Hagert (1985) stellt für jede Relation (links, rechts, über, unter) eigene Regeln zur Umkehr oder Konjunktion auf. Zur besseren Übersicht wurde in der Tabelle nur jeweils ein Beispiel genannt. Das „ \neg “ Zeichen bedeutet die Negation der Relation.

Die TML ist besonders aus historischer Sicht – durch ihre Nähe zur formalen Prädikatenlogik – sehr bedeutsam für das räumliche Schlussfolgern. Vor allem die neu aufkommende Möglichkeit, mentale Prozesse mit Computermodellen abzubilden, inspirierte viele Wissenschaftler. Trotzdem weist die TML beim räumlichen Schlussfolgern einige Schwächen auf und kann nicht alle Effekte ausreichend erklären. Sie stößt zum Beispiel an ihre Grenzen, wenn mehr als eine mögliche Anordnung aus den Prämissen entstehen kann (Byrne & Johnson-Laird, 1989). Deshalb herrscht mittlerweile weitgehend Konsens darüber, dass die Theorie Mentaler Modelle besser geeignet ist, um das räumliche Schlussfolgern abzubilden (Knauff & Knoblich, 2017).

1.2.1.2 Theorien mentaler Modelle

Die Theorien mentaler Modelle gehen auf die *semantische Wahrheitstheorie* zurück (Tarski, 1936). Demnach kann eine Konklusion nur wahr sein, wenn sie unter allen Interpretationen der Prämissen wahr ist. Das bedeutet, dass eine Konklusion nicht durch Regeln abgeleitet wird, sondern ihr Wahrheitswert durch die *Semantik* der Prämissen bestimmt wird (*Semantische Implikation*). Hier wird der größte Unterschied zu der mentalen Logik erkennbar. Für die TMM ist die syntaktische Struktur einer Prämisse für eine Inferenz nicht notwendig. Sie ist in der mentalen Repräsentation nicht einmal enthalten. Deshalb wird eine Konklusion nach diesen Theorien nicht formal abgeleitet, sondern auf *Validität* geprüft (Rips, 1994). Craik (1943) stellte die Hypothese auf, dass der Mensch – ähnlich zu einer Maschine – mit *Mentalen Modellen* denkt. Ein Mentales Modell nach Craik ist eine integrierte symbolische Repräsentation von in der Realität existierenden Gegebenheiten. Dabei müssen Symbole der *eindeutigen Determiniertheit* folgen. Das bedeutet, dass ein Symbol immer nur ein Objekt repräsentieren kann. Ein Modell muss die gleiche relationale Struktur wie die reale Gegebenheit aufweisen. Sie muss also physisch – in der Maschine als Subsystem und im Gehirn als neuronale Aktivität – repräsentiert sein. Das Denken mit Modellen läuft nach Craik in drei Schritten ab. Beim *Übersetzen* werden externe Prozesse in Worte, Zahlen, oder Symbole überführt. Dann folgt die *Deduktion*, bei der auf andere Symbole geschlussfolgert wird. Zuletzt wird die symbolische Repräsentation in externe Prozesse durch *Rückübersetzung* zurückgeführt. Craik nennt zur Veranschaulichung dieser Prozesse die Gezeitenmaschine nach Kelvin. Diese dient als Vorhersageinstrument von Ebbe und Flut. Die Maschine besteht aus vielen Subsystemen, welche sich aus Flaschenzügen und Getrieben zusammensetzen. Jedes dieser Subsysteme stellt eine Übersetzung eines der Einflussfaktoren (z. B. Sonne, Mond oder andere Planeten) auf die Gezeiten dar. Zusammen bilden diese Subsysteme das Modell der Gezeiten, aus deren Zusammenwirken (Deduktion) andere Getriebebewegungen entstehen können. Die daraus entstehenden Simulationen von Ebbe und Flut werden mit Hilfe eines am Modell befestigten Graphitstabes als oszillierende Linie auf ein Stück Papier übertragen (Rückübersetzung).

Auf Basis von Craiks Hypothese entwickelten Johnson-Laird (1983) und Johnson-Laird und Byrne (1991) ihre einflussreiche *Theorie Mentaler Modelle* zum deduktiven Denken. Die Grundannahme der Theorie ist, dass ein Schluss nur dann valide ist, wenn er unter allen möglichen Interpretationen (Modellen) der Prämissen gilt. Die TMM folgt dabei in ihrer ursprünglichen Form dem *Wahrheitsprinzip*. Es wird nur das in einem Modell repräsentiert, was wahr ist. Dies entspricht einer der logischen Definitionen von Modellen

(Knauff, 2006). Ein Modell ist auch nach der TMM eine reduzierte Repräsentation von realen Gegebenheiten und die TMM geht ebenfalls von drei Schritten der Verarbeitung von logischen Problemen aus. Zuerst kommt das *Verstehen* der Prämissen. Dabei nutzen die Personen ihr vorhandenes Wissen über Sprache und allgemeine Gegebenheiten. Sie erstellen eine interne Repräsentation von dem, was die Prämissen darstellen. Dabei gehen die verbalen Informationen in der Repräsentation verloren und bleiben nur implizit im Modell enthalten. Nach dem Verstehen kommt das *Beschreiben*. An dem zuvor gebildeten Modell wird eine vorläufige Konklusion abgelesen. Wenn nichts Neues abgelesen werden kann, folgert man, dass aus den Prämissen nichts folgt. Im letzten Schritt, dem *Validieren*, suchen Personen nach weiteren, alternativen Modellen, in denen ihre ursprüngliche Konklusion nicht hält. Wenn sich kein alternatives Modell konstruieren lässt, schlussfolgern sie, dass ihre Konklusion valide ist. Wenn sich ein solches Modell finden lässt, wird die zweite Phase wiederholt und versucht eine Konklusion zu finden, die unter allen Modellen hält. Darauf folgt wieder die Phase des Validierens. Wenn eine Prämisse zu mehreren Möglichkeiten führt, müssen mehrere Modelle gebildet werden, um korrekte Schlussfolgerungen ziehen zu können. Das Schlussfolgern nach der TMM mit einem drei-Term Problem sieht also folgendermaßen aus:

	Prämisse	Modell
1. Prämisse	Der Apfel ist links von der Birne.	Apfel Birne
2. Prämisse	Die Birne ist links von der Mango.	Apfel Birne Mango
Konklusion	Der Apfel ist links von der Mango.	Valide

Die TMM lässt einige Fragen bezüglich der genauen Form der Repräsentation offen. Deshalb hat Knauff (2013) die Definition der mentalen Modelle für das relationale Schließen spezifiziert. In seiner *Space to Reason Theorie* (SRT) postuliert er die *Spatial Layout Models*. Bei dieser besonderen Form der Repräsentation gehen verbale und bildliche Informationen aus den Prämissen gänzlich verloren. Übrig bleibt eine abstrakte integrierte Repräsentation, welche die Prämissen in rein räumlicher Form darstellt. Dies ist im Vergleich zur klassischen TMM ein bedeutender Unterschied. Die TMM macht ungenauere Angaben über den visuellen Anteil einer Repräsentation. Nach der SRT sind leicht zu visualisierende Relationen sogar schwerer zu verarbeiten. Dieser *visuelle Beeinträchtigungseffekt* (Knauff, 2013) konnte in mehreren Untersuchungen nachgewiesen werden (Gazzo Castañeda & Knauff, 2013; Hamburger, Ragni, Karimpur, Franzmeier, Wedell, & Knauff, 2018; Knauff & Johnson-Laird,

2002). Ein Layout nach der SRT kann man sich wie ein „Raster“ vorstellen, in dem ein Term in eine Zelle positioniert werden kann. Das Einfügen der Terme in das Layout geschieht nicht zufällig, sondern folgt dem Prinzip des *first free fits* (Ragni, Fangmeier, Webber, & Knauff, 2006; Ragni, Knauff, & Nebel, 2005). Ein Term wird also immer in das nächste freie Feld des Rasters eingefügt. Dabei ist die Richtung des Einfügens abhängig von den verbalen Eigenschaften der Prämisse. So werden bei der Prämisse „Der Apfel ist links von der Birne“ die beiden Terme als funktionell asymmetrisch angesehen (Talmy, 1983; Landau & Jackendoff, 1993; Hayward & Tarr, 1995). Der Apfel ist der Term, welcher in Relation zur Birne lokalisiert wird. Darum nennt man es auch das zu *lokalisierendes Objekt* (LO). Die Birne ist der Term, welcher als Referenz dient und wird deshalb als *Referenzobjekt* (RO) bezeichnet. Wenn durch das Einfügen der Terme mehrere Interpretationen möglich sind, müssen weitere Modelle berücksichtigt werden. Hier werden aber nicht mehrere Modelle gleichzeitig repräsentiert. Vielmehr wird ein Modell konstruiert und mit einer *Annotation* (mentale Fußnote, Rauh, 2000) versehen, dass eine Prämisse mehrere Interpretationen zulässt (Vandierendonck et al., 2004). Das heißt, dass nach der SRT immer nur *eine* Repräsentation gebildet und manipuliert wird (Knauff, 2013).

Die Annahme, dass Menschen nur eine Repräsentation bilden, hat weitreichende Folgen über das Format der Repräsentation hinaus. So sind die Prozesse beim Denken stark von der Repräsentation abhängig. Deshalb haben Knauff, Rauh, Schlieder und Strube (1998b) die drei Phasen der TMM für das räumliche Schlussfolgern präzisiert. So sprechen sie von den Phasen der *Modellkonstruktion*, *Modellinspektion*, und *Modellvariation*. Die wichtigsten Unterschiede zur TMM nach Johnson-Laird und Byrne (1991) liegen dabei in der ersten und der dritten Phase. Bei der Modellkonstruktion wird immer das kognitiv am leichtesten zu konstruierende Modell erstellt. Dieses *präferierte mentale Modell* (Jahn, Knauff, & Johnson-Laird, 2007; Knauff, Rauh, & Schlieder, 1995) ist eine bedeutsame Ergänzung der Theorie, da es viele Fehler beim räumlichen Schlussfolgern erklären kann. Die Autoren konnten zeigen, dass Personen dazu neigen, nach dem präferierten mentalen Modell die Suche nach Alternativen zu beenden und falsch antworten, weil sie die dritte Phase – die Variation – auslassen. Bei dieser Phase werden keine neuen Modelle konstruiert, sondern das ursprüngliche präferierte Modell wird variiert. Fehler entstehen also vor allem dadurch, dass Modelle falsch konstruiert werden oder bei mehreren möglichen Modellen die Annotation über weitere Interpretationen der Prämissen entweder nicht berücksichtigt oder vergessen wird (Rauh, Hagen, Knauff, Kuß, Schlieder, & Strube, 2005). Deshalb sind die Phasen der Konstruktion und Variation essenziell für das räumliche Schlussfolgern.

Allgemein ist ein Problem zum räumlichen Schlussfolgern umso schwieriger, desto mehr Modelle für eine valide Konklusion berücksichtigt werden müssen (Byrne & Johnson-Laird, 1989; Roberts, 2000). Es gibt jedoch auch weitere Faktoren, die die Schwierigkeit beim räumlichen Schlussfolgern beeinflussen. Kießner und Ragni (2017) zeigten in einer umfassenden Metaanalyse, dass Aufgaben mit geschriebenen Prämissen leichter zu lösen sind als gesprochene. Das Verifizieren – also Abgleichen einer vorgegebenen Konklusion mit dem eigenen Modell – fällt ebenfalls leichter als selbst eine oder mehrere Konklusionen zu generieren. Erklärt werden diese Effekte durch eine unterschiedlich starke Belastung des Arbeitsgedächtnisses (Baddeley, 1992). So müssen zum Beispiel bei auditiver Präsentation der Prämissen diese erst in der phonologischen Schleife gehalten und gleichzeitig im visuell-räumlichen Notizblock als Modell konstruiert werden.

Um die verschiedenen Phasen der SRT zu prüfen, nutzen Forscher unterschiedliche Aufgaben und damit assoziierte Effekte. Für die Phase der Konstruktion wurde der sogenannte *Kontinuitätseffekt* genutzt. Dieser zeigt, dass die Reihenfolge der Prämissen einen deutlichen Einfluss darauf hat, wie mentale Modelle konstruiert werden. Das bedeutet, dass sich allein durch das Tauschen der Prämissen die Verarbeitungsdauer und Fehlerrate erhöhen können (Knauff, Rauh, Schlieder, & Strube 1998a). Für die Variation hat sich der *Unbestimmtheitseffekt* als geeignetes Instrument herausgestellt. Bei diesem Effekt sind mehrere Modelle möglich und man ist für eine richtige Lösung des Problems gezwungen, das initial präferierte konstruierte Modell zu variieren. Der Kontinuitätseffekt und Unbestimmtheitseffekt stellten sich als sehr robuste Effekte beim räumlichen Schlussfolgern von Individuen heraus. Deshalb eignen sie sich ideal, um sie ebenfalls im kollaborativen Kontext mit Dyaden zu testen. Aus diesem Grund stelle ich im folgenden Abschnitt den Kontinuitätseffekt und den Unbestimmtheitseffekt mit den empirischen Befunden zum individuellen Denken dar.

1.2.2 Psychologische Befunde zum räumlichen Schlussfolgern

Um das räumliche Schlussfolgern genau zu untersuchen, nutzen Psychologen für ihre Experimente unterschiedliche Probleme. Die Aufgaben sind dazu konstruiert, die verschiedenen Phasen beim Denken mit räumlichen Inferenzaufgaben zu prüfen. Nach Johnson-Laird und Byrne (1991) sind klassische drei-Term Aufgaben jedoch oft zu leicht, um aussagekräftige Unterschiede zu finden. Je nach Anforderung wird deshalb der Grad der Komplexität einer Aufgabe beispielsweise durch Hinzufügen weiterer Terme oder Dimensionen erhöht (Halford, Wilson, & Phillips, 1989). Gemessen werden dann die

korrekten Konklusionen und die Zeit, die ein Proband zum Antworten benötigt. Außerdem erlauben die Lesezeiten der Prämissen Rückschlüsse auf den Konstruktionsprozess mentaler Modelle. Umso länger eine Prämisse gelesen wird, desto schwieriger ist sie zu verarbeiten. Diese Annahme ist Gegenstand des folgenden Effektes: des Kontinuitätseffektes.

1.2.2.1 Kontinuitätseffekt

Um die Phase der Modellkonstruktion zu untersuchen, werden vier-Term Probleme mit den Prämissen der Struktur (1) Ar_1B , (2) Br_2C und (3) Cr_3D verwendet. Dabei wird eine Prämisse nach der anderen präsentiert und die Versuchspersonen können selbstständig entscheiden, wann sie zu der nächsten Prämisse wechseln wollen. Nach der dritten Prämisse werden sie gebeten, das resultierende Modell (ABCD) zu generieren (Generierungsaufgabe). Die wichtigste Variation bei dieser Art von Problemen stellt die *Prämissenreihenfolge* dar:

Problem 1

- | | |
|---|--------------|
| (1) Der Apfel ist links von der Birne. (Ar_1B) | Modell: AB |
| (2) Die Birne ist links von der Mango. (Br_2C) | Modell: ABC |
| (3) Die Mango ist links von der Orange. (Cr_3D) | Modell: ABCD |

In diesem Beispiel baut die Reihenfolge der Terme aufeinander auf und das mentale Modell kann ohne weitere Probleme von links nach rechts konstruiert werden. Die Reihenfolge nennt man deshalb auch *kontinuierlich* (Ar_1B , Br_2C , Cr_3D). Das zweite Problem folgt einer anderen Struktur:

Problem 2

- | | |
|---|--------------|
| (1) Die Birne ist links von der Mango. (Br_2C) | Modell: BC |
| (2) Die Mango ist links von der Orange. (Cr_3D) | Modell: BCD |
| (3) Der Apfel ist links von der Birne. (Ar_1B) | Modell: ABCD |

Die ersten beiden Prämissen lassen sich wie beim ersten Beispiel als Modell von links nach rechts konstruieren. Der letzte Term muss jedoch nicht nach ganz rechts, sondern ganz links integriert werden. Deshalb nennt man diese Reihenfolge auch *semi-kontinuierlich* (Br_2C , Cr_3D , Ar_1B). Beim dritten Problem wird nun die Reihenfolge ein weiteres Mal geändert:

Problem 3

- | | |
|---|---------------|
| (1) Die Mango ist links von der Orange. (Cr_3D) | Modell: CD |
| (2) Der Apfel ist links von der Birne. (Ar_1B) | Modell: CD AB |
| (3) Die Birne ist links von der Mango. (Br_2C) | Modell: ABCD |

Hier ist es nicht möglich, ein aufeinander aufbauendes Modell zu konstruieren. Vielmehr müssen die ersten beiden Prämissen vorab repräsentiert werden und erst mit der letzten Prämisse kann eine integrierte Repräsentation konstruiert werden (Hunter, 1957). Deshalb nennt man diese Reihenfolge auch *diskontinuierlich* (Cr_3D , Ar_1B , Br_2C). Alle drei Beispiele führen zu dem gleichen Modell (Apfel Birne Mango Orange) und lassen die gleichen Konklusionen zu (Apfel r Mango, Apfel r Orange, Birne r Orange). Trotzdem konnte gezeigt werden, dass Probleme mit diskontinuierlicher Reihenfolge nicht nur am schwierigsten zu lösen sind, sondern vor allem die dritte Prämisse deutlich länger verarbeitet werden muss (Ehrlich & Johnson-Laird, 1982). Die Verarbeitungszeit bedeutet dabei die Dauer, die eine Person zum Lesen einer Prämisse benötigt.

Die Konstruktion des Modells hängt also von der Prämissenreihenfolge ab und man war daran interessiert, wie diese Art der Aufgaben repräsentiert und verarbeitet wird. So wäre eine Möglichkeit, zunächst die ersten beiden Prämissen als einzelne Modelle im Arbeitsgedächtnis zu halten und erst in ein Modell zu integrieren, wenn die dritte Prämisse präsentiert wird. Die SRT sagt jedoch vorher, dass Personen die beiden Prämissen als ein zusammenhängendes Modell repräsentieren, eine Annotation über die Diskontinuität im Arbeitsgedächtnis halten und anhand der Informationen aus der dritten Prämisse das Modell variieren. Um dies zu überprüfen, nutzten Nejasmic, Bucher und Knauff (2015) eine simple und doch geschickte Manipulation. Sie führten eine vierte Reihenfolge, die *quasi-diskontinuierlichen* Probleme ein:

- | | |
|---|---------------|
| (1) Die Mango ist links von der Orange. (Cr_3D) | Modell: CD |
| (2) Der Apfel ist links von der Birne. (Ar_1B) | Modell: CD AB |
| (3) Die Orange ist links von dem Apfel. (Dr_2A) | Modell: CDAB |

In diesem Fall erhält man auch erst durch die dritte Prämisse ausreichend Informationen über das zusammenhängende Modell. Wenn aber schon vorab die ersten zwei Prämissen als (vorläufig) zusammenhängendes Modell repräsentiert werden, ist mit der dritten Prämisse keine weitere Variation notwendig. Das bedeutet, dass diese Art der Aufgabe schneller lösbar ist als diskontinuierliche Probleme. Wenn aber die ersten beiden Prämissen als getrennte Modelle im Arbeitsgedächtnis gehalten werden und man diese Modelle erst mit der dritten Prämisse zusammenfügt, sollte es keinen Unterschied zwischen diskontinuierlichen und quasi-diskontinuierlichen Problemen geben. Die Autoren konnten zeigen, dass Personen bei diskontinuierlichen Problemen ganze zwei Sekunden länger für die Verarbeitung der dritten Prämisse benötigen als bei quasi-diskontinuierlichen Problemen. Außerdem waren die diskontinuierlichen Probleme signifikant schwieriger (76.8 % vs. 63.7 %) als quasi-diskontinuierliche Probleme. Dies stimmt mit der Vorhersage der SRT überein. Personen repräsentieren räumliche Relationen in einem Modell und variieren dieses Modell, sobald sie die dritte Prämisse präsentiert bekommen. Der Kontinuitätseffekt konnte ebenfalls für temporale Relationen (A ist während B, Knauff et al., 1998a), andere nicht räumliche Relationen (A ist neuer als B, Oakhill & Garnham, 1985) und beim Wegfinden (Sugimoto & Kusumi, 2014) nachgewiesen werden.

Der Kontinuitätseffekt stellt also den starken Einfluss der Prämissenreihenfolge auf das Konstruieren räumlicher Repräsentationen dar. Dabei konnte gezeigt werden, dass Menschen ein vorläufiges Modell konstruieren und dieses mit der dritten Prämisse variieren. Wir wissen aber bis heute nicht, ob dieser Effekt auch bei kollaborativen Gruppen auftritt und diese möglicherweise mehr korrekte Modelle konstruieren als Individuen. Außerdem haben wir gesehen, dass schon beim Kontinuitätseffekt die Variation eines Modells wichtig ist, um die korrekte endgültige Repräsentation zu bilden. Dieser Schritt wird noch wichtiger, wenn beim Schlussfolgern mehr als eine eindeutige Interpretation der Prämissen möglich ist. Dies ist beim Unbestimmtheitseffekt der Fall.

1.2.2.2 Unbestimmtheitseffekt

Für die Untersuchung der Modellvariation beim räumlichen Schlussfolgern wurden zweidimensionale Probleme verwendet. Diese erhöhen den Grad der Komplexität (Halford et al., 1998) und ermöglichen folgende Arten von Problemen:

Problem 1

- (1) Der Apfel ist links von der Birne. (Ar₁B)
- (2) Die Birne ist links von der Mango. (Br₂C)
- (3) Die Orange ist unter dem Apfel. (Dr₄A)
- (4) Die Kiwi ist unter der Mango. (Er₅C)

Resultierendes Modell

Apfel	Birne	Mango
Orange		Kiwi

Bei diesem Problem wird immer nach der Relation der beiden unteren Terme (Orange r Kiwi) gefragt. In diesem Fall kann nur ein Modell konstruiert werden und somit ist eine klare Konklusion „Die Orange ist links von der Kiwi“ möglich, weshalb man das Problem ein *ein-Modell Problem* nennt. Wenn man aber eine leichte Änderung in der zweiten Prämisse vornimmt ergibt sich folgendes Problem:

Problem 2

- (1) Der Apfel ist links von der Birne. (Ar₁B)
- (2) Die Mango ist rechts von dem Apfel. (Cr₂A)
- (3) Die Orange ist unter dem Apfel. (Dr₄A)
- (4) Die Kiwi ist unter der Mango. (Er₅C)

Resultierende Modelle

Apfel	Birne	Mango	Apfel	Mango	Birne
Orange		Kiwi	Orange	Kiwi	

Durch diese Änderung sind zwei Modelle möglich. Die Relation von Orange und Kiwi ist jedoch bei beiden Modellen identisch. Darum bezeichnet man dieses Problem auch als ein *zwei-Modelle determiniertes Problem*. Da bei diesem Problem mehr als ein Modell möglich

ist, sollte diese Art von Problem nach der TMM und SRT schwieriger sein als ein Problem mit nur einem Modell. Theoretisch müsste für die korrekte Konklusion jedoch nur ein einziges (präferiertes) Modell konstruiert werden und man würde richtig schlussfolgern, dass die „Orange links von der Kiwi“ ist. Eine weitere Änderung der zweiten Prämisse führt zu folgendem Problem:

Problem 3

- (1) Der Apfel ist links von der Birne. (Ar₁B)
- (2) Die Mango ist links von der Birne. (Cr₂B)
- (3) Die Orange ist unter dem Apfel. (Dr₄A)
- (4) Die Kiwi ist unter der Mango. (Er₅C)

Resultierende Modelle

Apfel	Mango	Birne	Mango	Apfel	Birne
Orange	Kiwi		Kiwi	Orange	

Bei diesem Problem führen die Prämissen ebenfalls zu zwei möglichen Modellen. In diesem Fall ist die Relation „Orange r Kiwi“ nicht bestimmbar. Nach einem Modell wäre die Relation „links“ und die andere führt zu „rechts“. Deshalb kann man nur schließen, dass *keine valide Konklusion* möglich ist. Dieses Problem nennt man *zwei-Modelle indeterminiertes Problem*. Man kann diese Art von Problemen nur korrekt lösen, wenn beide Modelle berücksichtigt werden. Deshalb sollte dieses Problem nach der TMM am schwierigsten zu lösen sein. Dieser *Unbestimmtheitseffekt* wurde in vielen Experimenten zum Schlussfolgern untersucht. Das bekannteste Experiment stammt dabei von Byrne und Johnson-Laird (1989). Sie konnten nicht nur zeigen, dass mehr-Modell Probleme schwieriger als ein-Modell Probleme sind, sondern auch, dass indeterminierte Probleme am schwierigsten zu lösen sind. Die Probanden zogen bei ein-Modell Problemen im Durchschnitt 61 % korrekte Konklusionen. zwei-Modell determinierte Probleme wurden mit 50 % noch ähnlich häufig korrekt beantwortet. Bei indeterminierten zwei-Modell Problemen konnten die Probanden hingegen in nur 18 % der Fälle korrekt angeben, dass keine valide Konklusion möglich ist. Carreiras and Santamaria (1997) konnten zeigen, dass sich die Schwierigkeit ebenfalls in der Verarbeitungsdauer niederschlägt. Indeterminierte Probleme benötigen deutlich mehr

Verarbeitungszeit als determinierte. Ebenfalls benötigen auch determinierte mehr-Modell Probleme mehr Zeit, als ein-Modell Probleme. Roberts (2000) konnte zudem zeigen, dass die Art der Darbietung von Prämissen eine Rolle dabei spielt, ob Personen mehr als ein Modell berücksichtigen. So profitieren Menschen davon, wenn die Prämissen parallel – also alle zusammen auf einem Bild – präsentiert werden. Dies führt zur Entlastung des Arbeitsgedächtnisses (Baddeley, 1992) und erhöht die Wahrscheinlichkeit, dass wir die Annotation über weitere Interpretationen der Prämissen berücksichtigen. Dass das Arbeitsgedächtnis eine wichtige Rolle beim Variieren mentaler Modelle spielt, konnten auch Vandierendonck und De Vooght (1997) zeigen. Sie ließen die Probanden räumliche oder temporale Inferenzaufgaben lösen. Das Besondere dabei war, dass die Probanden gleichzeitig eine Zweitaufgabe lösen mussten. Diese Aufgaben waren so konstruiert, dass sie unterschiedliche Komponenten des Arbeitsgedächtnisses belasteten. Ihre Befunde deuteten darauf hin, dass vor allem die visuell-räumliche Komponente des Arbeitsgedächtnisses für das Verarbeiten mentaler Modelle zuständig ist.

Mit den zahlreichen Untersuchungen zum Unbestimmtheitseffekt konnten die Forscher zeigen, wie wichtig die Phase der Modellvariation ist, um korrekte Schlussfolgerungen bei indeterminierten Problemen zu generieren. Dabei ist es wichtig, dass – nachdem zunächst das präferierte Modell konstruiert wurde – die Annotation über weitere Interpretationen der Prämissen berücksichtigt wird. Nur dann kann man damit beginnen, das präferierte Modell zu variieren (Rauh et al., 2005). Wir haben gesehen, dass diese Aufgabe sehr schwierig ist und überwiegend auf die hohe Belastung für die räumliche Komponente des Arbeitsgedächtnisses zurückzuführen ist. Eine besondere Form der Variation tritt beim Schlussfolgern auf, wenn wir bereits eine integrierte Repräsentation gebildet haben und plötzlich mit neuer – zu der gebildeten Überzeugung inkonsistenter – Information konfrontiert werden. In diesem Fall muss die ursprüngliche Überzeugung revidiert werden. Die Revision von Überzeugungen ist stark mit den bereits vorgestellten Prozessen zum Schlussfolgern verknüpft und soll im folgenden Abschnitt genauer erläutert werden.

1.3 Überzeugungsrevision

Stellen Sie sich vor, Sie sind mit einem Freund in einem Café verabredet. Sie kennen das Café nicht und Ihr Freund sagt Ihnen, dass Sie vom Bahnhof zum Einkaufszentrum laufen sollen. Das Café finden Sie dann *links* von dem Einkaufszentrum. Sie gehen also zum Einkaufszentrum, aber können nur ein Café *rechts* davon entdecken. Sie können dafür verschiedene Erklärungen finden. Zum Beispiel könnten könnte Ihr Freund die Richtungen

vertauscht haben (z. B. Bucher, Röser, Nejasmic, & Hamburger, 2014). Oder Sie stehen auf der falschen Seite des Einkaufszentrums. Sicher ist, dass Ihre ursprüngliche Überzeugung (das Café ist links von dem Einkaufszentrum) nicht richtig sein kann. Durch solche Widersprüche können wir die neue Information nicht einfach unserem Wissenssystem hinzufügen und sind gezwungen, zumindest einen Teil unserer ursprünglichen Überzeugung zu revidieren (Katsuno & Mendelson, 1991). Diese Form des Denkens wird als *Überzeugungsrevision* bezeichnet und ist besonders stark in der Sozialpsychologie von Bedeutung. So ist die Überzeugungsrevision beispielsweise ein wichtiger Prozess bei der Änderung von Einstellungen (Vogel & Wanke, 2016). Ziel der Überzeugungsrevision ist es, kognitive Konflikte aufzudecken und Konsistenz wiederherzustellen (Gawronski & Strack, 2012). Erkenntnistheoretisch wird die Überzeugungsrevision als „zeitliche Veränderung epistemischer Zustände unter dem Einfluss neuer Informationen“ verstanden (Haas, 2019, S. 354). Beim Schlussfolgern bedeutet dies, dass eine zuvor gezogene Konklusion auf Grund eines unanfechtbaren Faktes revidiert werden muss. Deshalb weist die Überzeugungsrevision keine *Monotonie* auf (Elio & Pelletier, 1997). Unter Monotonie versteht man, dass zusätzliche Informationen vorhergegangene Schlussfolgerungen nicht ungültig machen. Sie lassen lediglich weitere Schlussfolgerungen zu. Bei der Überzeugungsrevision ist dies anders. Ein unanfechtbarer Fakt führt dazu, dass ursprüngliches Wissen verworfen werden muss. Die Erforschung solcher Überzeugungsrevisionen aus Sicht der Kognitionswissenschaften wurde durch die Theorien zur künstlichen Intelligenz (KI) geprägt (Knauff & Knoblich, 2017). Harman (1986) und Rott (2001) definieren verschiedene Prinzipien, denen ein intelligentes System beim Revidieren folgen sollte. Das Prinzip des *Konservatismus* besagt, dass eine Überzeugung solange beibehalten wird, bis neue (inkonsistente) Informationen eine Revision erfordern. Nach dem *Prinzip epistemischer Verwurzelung* soll nur das Wissen verworfen werden, welches am wenigsten grundlegend für ein System ist. Man geht also davon aus, dass bestimmtes Wissen stärker begründet ist und deshalb weniger häufig verworfen wird (Knauff & Knoblich, 2017). Formal bedeutet dies, dass das Prinzip über die *Verwurzelungsrelation* ($A > B$) von zwei (oder mehr) Überzeugungen durchgeführt wird. Diese Relation erfüllt dabei auch die Eigenschaft der Transitivität (Makinson & Gärdenfors, 1988). Man kann also aus „A ist stärker verwurzelt als B“ und „B ist stärker verwurzelt als C“ automatisch „A ist stärker verwurzelt als C“ schlussfolgern und würde somit „C“ als erstes und „A“ als letztes aufgeben. An der Verwurzelungsrelation kann man also sehen, wie nah die Überzeugungsrevision den Prozessen des Schlussfolgerns ist. Ein weiteres Prinzip ist das der *minimalen Veränderung*. Es besagt, dass man nur so wenig wie möglich an der ursprünglichen Überzeugungsbasis

verändern sollte. Damit ist gemeint, dass die Änderung und ihre Auswirkung auf weitere Überzeugungen minimal gehalten werden. Wenn man durch eine Revision eine oder zwei Prämissen verwerfen könnte, sollte die Revision, die nur eine Prämisse beeinflusst, gewählt werden.

Ein formales System zur Überzeugungsrevision, das diese Prinzipien enthält, stellt das nach ihren Autoren benannte AGM-System dar (Alchourrón, Gärdenfors, & Makinson, 1985). Dieses System ist noch heute eine wichtige Basis um zu beschreiben, wie künstliche Systeme mit Inkonsistenzen umgehen sollen. So sollen bei einer Revision nicht alle Überzeugungen gelöscht werden. Wie beim Prinzip der minimalen Veränderung sollen nur die Überzeugungen verworfen werden, die für eine Revision notwendig sind. Außerdem sollen alle Informationen, die sich aus der neuen Überzeugung ableiten lassen, ebenfalls in die Wissensbasis übernommen werden. Umgesetzt werden diese Regeln in drei Schritten. Bei der *Expansion* wird das vorhandene Überzeugungssystem um neue (inkonsistente) Informationen erweitert. Der Prozess der *Revision* soll diese Überzeugungen in ein konsistentes System überführen. Auch bei der Revision wird das Überzeugungssystem um neue Überzeugungen erweitert. Falls es dabei zu Redundanzen kommt, müssen diese im letzten Schritt – der *Kontraktion* – eliminiert werden (Alchourrón et al., 1985; Fermé & Hansson, 2011). Dabei werden Redundanzen nicht zufällig eliminiert, sondern folgen wie bei der Revision wieder dem Prinzip der epistemischen Verwurzelung (Makinson & Gärdenfors, 1988; Rott, 1992). Das AGM-System und die Prinzipien zur Überzeugungsrevision sind Basis vieler psychologischer Experimente. Dabei interessiert die Forscher besonders, ob und wie die Prinzipien der epistemischen Verwurzelung und das Prinzip der minimalen Veränderung vom Menschen realisiert werden (Elio & Pelletier, 1997).

1.3.1 Psychologische Befunde zur Überzeugungsrevision

Allgemein gehen die Vertreter der KI-Systeme davon aus, dass Menschen dem Prinzip der minimalen Veränderung bei der Revision folgen. Psychologische Ansätze sehen eher die epistemische Verwurzelung als ausschlaggebend. Um Letzteres zu testen, präsentierten Revlis, Lipkin und Hayes (1971) den Probanden Syllogismen mit folgender Struktur:

- (1) Alle Menschen sind intelligente Wesen.
- (2) Jens ist ein Mensch.

Fakt: Jens ist kein intelligentes Wesen.

Die erste Aussage gilt als allgemeine Regel und die zweite als spezielle Aussage. Der Fakt ist dabei unanfechtbar wahr, aber nicht mit den vorherigen Aussagen vereinbar. Man muss sich also entscheiden, eine der beiden Aussagen aufzugeben. Probanden entschieden sich in 73 % der Fälle dafür, die zweite Aussage zu verwerfen. Das bedeutet, dass sie eher bereit waren zu akzeptieren, dass Jens kein Mensch ist anstatt allen Menschen die Intelligenz abzuerkennen, indem sie die erste Aussage ablehnen würden. Sie verwerfen also eher die zweite (spezielle) Aussage. Dies entspricht exakt dem Prinzip der epistemischen Verwurzelung. Die Autoren konnten zusätzlich zeigen, dass Menschen dabei nicht nur die Form mit der Verwurzelungsrelation „(1) > (2)“, sondern auch den Inhalt berücksichtigen. So entschieden sie sich in 76 % der Fälle, eine für sie bekannte, allgemeine Aussage beizubehalten. War diese Aussage ohne sinnvollen Inhalt versehen (z. B. Alle Biber sind blau. Jan ist ein Biber. Fakt: Jan ist nicht blau), sank diese Zustimmung auf 70 %.

Elio und Pelletier (1997) testeten das Prinzip epistemischer Verwurzelung mit konditionalen Aussagen:

- | | | |
|--------------|--|--------------|
| (1) | Wenn es regnet, dann wird die Straße nass. ($P \rightarrow Q$) | |
| (2) | Es regnet. | (P) |
| Fakt: | Die Straße wird nicht nass. | ($\neg Q$) |

Die erste Aussage – das Konditional ($P \rightarrow Q$) – gilt wie beim vorherigen Beispiel als die allgemeinere Aussage, da sie die Regel enthält. Die Autoren kamen zu überraschenden Ergebnissen. Anders als bei Revlis et al. (1971) entschieden sich die Probanden in 75 % der Fälle dazu, die erste – allgemeine – Aussage zu verwerfen und somit nicht dem Prinzip der epistemischen Verwurzelung zu folgen. Die Erklärung der Autoren basiert auf einer empiristischen Sicht des Denkens. So stellt die zweite (spezielle) Prämisse einen Datenpunkt dar. Die erste Prämisse stellt eine Regel dar, die aus den Daten entstanden ist. Der neue Fakt ist ein weiterer Datenpunkt der lediglich mit der Regel, nicht aber mit der zweiten Prämisse im Konflikt steht. Deshalb wird die erste Aussage eher verworfen. Diese Annahme wird durch weitere Befunde von Elio und Pelletier (1997) gestützt, bei denen der Effekt stärker wurde, wenn man die Aufgaben mit (fiktiven) Inhalten präsentierte. In weiteren Experimenten testeten die Autoren, ob das Prinzip minimaler Veränderung die Überzeugungsrevision besser erklären kann. Auch hier kamen sie zu dem Schluss, dass Personen keinem eindeutigen Prinzip folgen.

Johnson-Laird, Girotto und Legrenzi (2004) versuchten die Überzeugungsrevision aus einer anderen Perspektive zu betrachten und postulieren, dass nicht die syntaktische Struktur, sondern die Semantik entscheidet, welche Aussage man verwirft. Das heißt, dass die Prämisse verworfen wird, die ausschließlich zu Modellen führt, die nicht mit dem Fakt vereinbar sind. Daher wird dieses Prinzip auch das *mismatch principle* genannt. Gibt es keine Prämisse, die ausschließlich zu dem Fakt widersprüchlicher Modelle führt, wird die Prämisse verworfen, die zu mindestens einem widersprüchlichen Modell führt oder den Fakt nicht repräsentiert. Kann keines solcher Modelle gefunden werden, wird die zu verwerfende Prämisse mit gleicher Wahrscheinlichkeit ausgewählt. Zum Beispiel:

- (1) Wenn es regnet, dann wird die Straße nass. ($P \rightarrow Q$)
 (2) Es regnet. (P)
Fakt: Die Straße wird nicht nass. ($\neg Q$)

In diesem Fall widerspricht das Modell des Faktes ($\neg Q$) lediglich dem Modell der ersten Aussage ($P \rightarrow Q$), nicht aber der zweiten Aussage (P) und die erste Aussage wird verworfen. Ein ähnliches Problem lautet:

- (1) Wenn es regnet, dann wird die Straße nass. ($P \rightarrow Q$)
 (2) Die Straße wird nicht nass. ($\neg Q$)
Fakt: Es regnet. (P)

In diesem Fall wird der Fakt (P) durch die erste Aussage ($P \rightarrow Q$) repräsentiert, nicht aber durch die zweite ($\neg Q$). Deshalb verwerfen Personen nach dem *mismatch principle* in diesem Fall die zweite Prämisse häufiger.

Hasson und Johnson-Laird (2003) testeten das *mismatch principle*, indem sie die Aussagen in einen Gesprächskontext einbetteten und die Probanden beurteilen ließen, wem sie mehr glauben. Zusätzlich variierten sie, ob lediglich zwei Aussagen oder ein dritter unanfechtbarer Fakt präsentiert wurden:

(1) Person A: Wenn das Experiment richtig durchgeführt wurde,
dann ist das Helium flüssig. $(P \rightarrow Q)$

(2) Person B: Das Helium ist nicht flüssig. $(\neg Q)$

In den meisten Fällen entschieden sich die Probanden, der Person B (und deren Aussage) mehr zu glauben. Wenn jedoch ein zusätzlicher Fakt „Das Experiment wurde richtig durchgeführt“ (P) präsentiert wurde, entschieden sie sich plötzlich, die zweite Aussage zu verwerfen. Dieses Ergebnis hat weitreichende Folgen für die epistemische Verwurzelung. Würde man die Revision ausschließlich anhand stärker verwurzelter (glaubwürdiger) Aussagen machen, sollte der Fakt keinen Einfluss darauf haben, welcher Person man mehr glaubt. Die Autoren zeigten jedoch, dass genau dieser Fakt die Revision bestimmt. Der Fakt führt zu mit den Aussagen inkonsistenten Modellen und somit bestimmt nicht die Glaubwürdigkeit der Aussage die Revision, sondern das *mismatch principle*.

Hasson und Johnson-Laird (2003) konnten also zeigen, dass in Gesprächskontexten, bei denen die Glaubwürdigkeit der Person über deren Aussage bestimmt wird, das *mismatch principle* die Revisionsentscheidungen erklären kann. Doch was passiert, wenn man die Glaubwürdigkeit der Quelle direkt beeinflusst? Wolf, Rieger und Knauff (2012) präsentierten den Probanden Konditionale, die ebenfalls in einen Unterhaltungskontext eingebettet waren. Die Unterhaltung fand dabei immer zwischen zwei Personen aus verschiedenen Berufsgruppen statt. Die Vertrauenswürdigkeit dieser Berufe wurde vorab in einer anderen Studie ermittelt. So konnten die Autoren die unterschiedlich glaubwürdigen Berufe in der Unterhaltung gezielt manipulieren. Eine Unterhaltung zwischen einem Autohändler (wenig vertrauenswürdig) und einem Polizisten (sehr vertrauenswürdig) wurde folgendermaßen präsentiert:

(1) Autohändler: Wenn A der Fall ist, dann ist B der Fall. $(P \rightarrow Q)$

(2) Polizist: B ist nicht der Fall. $(\neg Q)$

Fakt: A ist der Fall. (P)

Die Aufgabe war wieder, den unanfechtbaren Fakt zu berücksichtigen und eine der Aussagen zu verwerfen. Die Ergebnisse zeigen, dass weder das Prinzip epistemischer Verwurzelung, noch das *mismatch principle* die Entscheidung leiteten. Die Probanden folgten dem sogenannten *Quellenprinzip*. Das bedeutet, dass die Vertrauenswürdigkeit der Quelle

ausschlaggebend dafür war, welche Aussage sie verwerfen. In diesem Beispiel würde man also die Aussage des Autohändlers verwerfen und die des Polizisten beibehalten.

In einem weiteren Experiment wollten die Autoren die Konversation in einen natürlicheren Kontext einbetten. Dafür passten sie den Inhalt der Aussagen an die Berufe an:

- (1) Polizist: Wenn Karl ein beschädigtes Auto verkauft,
dann bietet er es zu einem reduzierten Preis an.
 $(P \rightarrow Q)$
- (2) Autohändler: Karl verkauft ein beschädigtes Auto.
 (P)
- Fakt:** **Karl bietet es nicht zu einem reduzierten Preis an.**
 $(\neg Q)$

Wieder sollte man davon ausgehen, dass durch das Quellenprinzip der Aussage des Polizisten mehr Vertrauen zugesprochen wird und deshalb die zweite Aussage verworfen wird. Interessanterweise zeigte sich dieses Mal jedoch keine solche Präferenz. Dadurch schlussfolgerten die Autoren, dass der Inhalt der Aussagen mit der Vertrauenswürdigkeit der Quelle interagiert und den Revisionsprozess beeinflusst. Deshalb führten die Autoren ein drittes Experiment durch. Hier wurden die Aussagen inhaltsreich, aber nicht auf Personen bezogen, konstruiert. Außerdem nutzten die Autoren keine Berufsgruppen, sondern bildeten fiktive Beschreibungen über die Persönlichkeit von zwei Personen (Mark und Christian). Christian wurde als zuverlässig, verantwortungsvoll und vertrauenswürdig beschrieben. Mark hingegen wurden genau gegenteilige Attribute zugeschrieben:

- (1) Christian: Wenn Katharina am Abend kocht,
dann bereitet sie Reis zu. $(P \rightarrow Q)$
- (2) Mark: Katharina kocht am Abend. (P)
- Fakt:** **Katharina bereitet keinen Reis zu.** $(\neg Q)$

In diesem Fall entschieden sich die Probanden wieder, häufiger die Aussage von Mark zu verwerfen. Daraus schlossen die Autoren, dass nicht Inhalt per se das Quellenprinzip

beeinflusst, sondern nur dann, wenn man Hintergrundwissen über die Personen (Berufe) mit den Aussagen vergleichen kann.

Die Revision von Überzeugungen scheint also mehr als nur simplen formalen Prinzipien zu folgen. Dies macht die Untersuchung der Überzeugungsrevision auf der einen Seite zu einem sehr spannenden Thema. So ist offensichtlich keine direkte Interaktion mit einem anderen Diskussionspartner notwendig, um die Überzeugungsrevision zu beeinflussen. Das reine ‚Beobachten‘ einer Diskussion zwischen zwei Personen führt dazu, dass man dem Quellenprinzip folgt. Auf der anderen Seite führen die insgesamt sehr heterogenen Ergebnisse zu dem Problem, dass man nur schwer konsistente Vorhersagen tätigen kann. Robustere Befunde konnten jedoch bei Überzeugungsrevisionen mit räumlichen Inhalten gefunden werden. Diese zeigen erneut, dass die Phase der Modellvariation der ausschlaggebende Prozess beim Verarbeiten der Probleme ist.

1.3.2 Räumliche Überzeugungsrevision

Die Untersuchung räumlicher Überzeugungsrevision ist stark an die bereits vorgestellten Paradigmen angelehnt. Die Prämissen folgen der Struktur von vier Figuren:

	Figur 1	Figur 2	Figur 3	Figur 4
1. Prämisse	A links B	A links B	B rechts A	B rechts A
2. Prämisse	B links C	C rechts B	C rechts B	B links C

Jede dieser Figuren führt zu dem Modell „A B C“. Jedoch unterscheiden sie sich in der Art, wie sie konstruiert werden. So kann man gewährleisten, dass eine Revision unabhängig vom Konstruktionsprozess stattfindet. Zusätzlich wird ein unanfechtbarer Fakt mit dem Prädikat „links von“ oder „rechts von“ präsentiert. Ein Beispiel lautet:

- (1) Der Apfel ist links von der Birne
- (2) Die Birne ist links von der Mango

Korrektes Modell: Apfel Birne Mango

Fakt: Der Apfel ist rechts von der Mango.

Die Annahme ist, dass Personen zuerst versuchen, ein Modell zu konstruieren, das alle Prämissen inklusive des Faktes berücksichtigt. Ist dies möglich, ist das Modell *konsistent* und

keine weitere Operation ist notwendig. Ist dies jedoch wie in dem Beispiel nicht möglich, ist das Modell *inkonsistent* und eine Revision muss durchgeführt werden (Johnson-Laird et al., 2004). Der Prozess der Revision funktioniert – ähnlich wie beim räumlichen Schlussfolgern – durch Variation des ursprünglichen Modells (Apfel Birne Mango). Das bedeutet, dass keine neuen Modelle von Grund auf konstruiert werden, sondern anhand eines präferierten Modells variiert werden (Knauff, Bucher, Krumnack, & Nejasmic, 2013). Folgende Variationen des ursprünglichen Modells sind in dem vorliegenden Beispiel unter Berücksichtigung des Faktors möglich:

Birne Mango Apfel

oder

Mango Birne Apfel

Wählt man das erste Modell, wird die erste Prämisse (Der Apfel ist links von der Birne) verworfen. Wählt man das zweite, wird die zweite Prämisse (Die Birne ist links von der Mango) verworfen. Beide Modelle verwerfen also jeweils eine Prämisse unter Berücksichtigung des Faktors und sind somit gleichwertig korrekt. Es macht logisch keinen Unterschied, ob man sich für das erste oder zweite Modell entscheidet. Deshalb sagt auch das *mismatch principle* hier lediglich vorher, dass es keine Präferenz bei der Revision geben sollte (Johnson-Laird et al., 2004).

Doch wenn beide Modelle logisch gleichwertig sind und das *mismatch principle* keine Vorhersagen trifft, wie entscheiden wir dann, welche Prämisse wir verwerfen? Die psychologischen Untersuchungen zeigen, dass es eine klare Präferenz gibt. Die Revision hängt von der gleichen funktionellen Asymmetrie der Terme ab, die sich schon bei der Modellkonstruktion zeigte (Hayward & Tarr, 1995; Landau & Jackendoff, 1993; Talmy, 1983). So wird häufiger der Apfel in Relation zur Mango positioniert. Das heißt, dass wie beim Einfügen der Terme in das *spatial layout* bei dem Fakt „der Apfel ist rechts von der Mango“ der Apfel als flexibler und somit als *LO* und die Orange als *RO* verstanden wird (Landau & Jackendoff, 1993). Dieses *LO-Prinzip* stellte sich als sehr robuster Effekt heraus. Dabei revidieren die Probanden das *LO* nicht nur deutlich häufiger, sondern benötigen dafür auch deutlich weniger Zeit, als wenn sie das *RO* relokalisieren (Bucher, Krumnack, Nejasmic, & Knauff, 2011; Knauff et al., 2013).

Die Entdeckung des *LO-Prinzips* führte zu einer ganzen Reihe weiterer Untersuchungen. So konnten Bucher, Nejasmic, Bertleff und Knauff (2013) zeigen, dass das *LO-Prinzip* auch dann verwendet wird, wenn sie zu unplausiblen Konklusionen wie „Die Gurke ist dicker als der Kürbis“ führt. Dies ist ein Hinweis darauf, dass das *LO-Prinzip* nicht

durch die epistemische Verwurzelung beeinflusst wird. Im gleichen Experiment konnten Bucher et al. (2013) zeigen, dass eine Revision länger dauerte, wenn die Aufgabe leicht zu visualisieren war. Das bedeutet, dass Modelle, die mit leicht zu visualisierenden Prädikaten, wie „dreckiger“, konstruiert wurden, zwar auch durch das LO-Prinzip revidiert werden, jedoch länger brauchen als bei schwer zu visualisierenden Prädikaten wie „intelligenter“. Dies entspricht dem visuellen Beeinträchtigungseffekt (Knauff & Johnson-Laird, 2002), der bereits aus der Forschung zum räumlichen Schlussfolgern bekannt ist (Gazzo Castaneda & Knauff, 2013; Hamburger et al., 2018; Knauff & Johnson-Laird, 2002).

In einer anderen Studie haben Bucher und Nejasmic (2012) überprüft, ob das Prinzip minimaler Veränderung das LO-Prinzip beeinflusst. Sie variierten die Anzahl an Termen, die mit einer LO-Revision relokalisiert werden mussten. Sie konnten zeigen, dass entgegen dem Prinzip der minimalen Veränderung das LO-Prinzip sogar dann genutzt wurde, wenn dadurch im Vergleich zum RO-Prinzip mehr Objekte relokalisiert werden mussten. Das Prinzip minimaler Veränderung ist jedoch für das räumliche Schlussfolgern nicht genau definiert. Es kann als die Anzahl der zu verwerfenden Prämissen, die Anzahl zu bewegender Terme oder der kognitiven Schritte, die für eine Revision nötig sind, verstanden werden (Krumnack, Bucher, Nejasmic, & Knauff, 2011). Bucher und Thorn (2014) definierten die minimale Veränderung anhand der Anzahl der Prämissen, die verworfen werden müssen. Dazu nutzten sie drei-Term Aufgaben und fügten den Relationen den Zusatz „direkt“ ($A \text{ r}_{\text{direkt}} B$) hinzu. Dies lässt Aufgaben zu, bei denen man durch das LO-Prinzip beide ursprünglichen Prämissen verwerfen muss. Die Autoren konnten zeigen, dass Personen in diesem Fall nicht bereit waren, beide Prämissen zugunsten des LO-Prinzips aufzugeben. Deshalb entschieden sie sich häufiger das Referenzobjekt (Mango) zu relokalisieren und somit zumindest eine Prämisse beizubehalten.

Nejasmic, Bucher und Thorn (2014) verwendeten den Kontinuitätseffekt, um zu testen, ob der kognitive Aufwand beim Konstruieren der mentalen Modelle das LO-Prinzip beeinflusst. Dabei wurden die Prämissen in kontinuierlicher, semi-kontinuierlicher oder diskontinuierlicher Reihenfolge präsentiert. Die Prämissen selbst enthielten dabei ausschließlich das Prädikat „links von“. Beim Fakt variierten sie die Richtung, in die das LO revidiert werden musste (D ist links von A vs. A ist rechts von D). Interessant bei ihren Ergebnissen ist, dass das LO-Prinzip weiterhin vorherrscht, wenn konsistent eine (kontinuierliche, semi-kontinuierliche, oder diskontinuierliche) Prämissenreihenfolge präsentiert wurde (Experiment 1). Wenn jedoch die Prämissenreihenfolge als Innersubjektfaktor zufällig für jede Aufgabe gewählt wurde, war die Konstruktionsphase

kognitiv deutlich belastender. Dies führte dazu, dass die Revisionsstrategie nicht mehr ausschließlich dem LO-Prinzip folgte, sondern von der Konstruktion abhängig war: Personen nutzten häufiger die LO-Revision, wenn diese dem Konstruktionsprozess nach links entsprach. Führte das LO-Prinzip jedoch dazu, dass das Objekt nach rechts revidiert werden musste (entgegen der Konstruktionsrichtung), sank die Entscheidung auf 58 %. Daraus konnten die Autoren schließen, dass die Konstruktion einen Einfluss auf den Revisionsprozess hat. Das heißt, dass die Phase der Konstruktion eine Art mentales Set (Luchins & Luchins, 1959) für die Aufgabe definiert, anhand dessen Menschen den weiteren Schritt der Revision orientieren.

Die Untersuchungen konnten zeigen, dass die Revision räumlicher Überzeugungen auf der Variation eines zuvor konstruierten Modells basiert. Dabei geschieht die Revision bei zwei gleichwertigen Alternativen nicht zufällig, sondern ist sehr stark von der funktionellen Asymmetrie der Terme im Fakt abhängig. Dieses LO-Prinzip stellte sich als sehr robust heraus und erst durch besondere Prämissen (z. B. direkt links) oder komplexere Konstruktionsprozesse konnte man dieses Prinzip beeinflussen. Beim Denken mit nicht räumlichen Überzeugungen konnte gezeigt werden, dass die Revision durch das Quellenprinzip beeinflusst wird. Das bedeutet, dass nicht nur formale Regeln wie die epistemische Verwurzelung oder modelltheoretische Prozesse wie das *mismatch principle* einen Einfluss darauf haben, wie Menschen ihre Überzeugungen ändern. Vielmehr kann es ausreichen, wenn einer Person weniger Vertrauen zugesprochen wird und man sich deshalb nicht an formalen, sondern sozialen Prinzipien orientiert. Dieses Prinzip wurde bisher jedoch noch nicht bei der räumlichen Überzeugungsrevision getestet. Es muss erst noch beantwortet werden, ob Faktoren wie das Quellenprinzip oder andere soziale Einflüsse wie das Denken in Gruppen die Revision räumlicher Überzeugungen verändern können.

Im bisherigen Abschnitt habe ich gezeigt, dass das räumliche Schlussfolgern und die Revision räumlicher Überzeugungen mit den Phasen der Modellkonstruktion, Inspektion und Variation beschrieben werden können. Die Experimente zeigten starke Einflüsse durch die Reihenfolge der Prämissen oder die Anzahl der zu konstruierenden Modelle auf die Leistung beim Schlussfolgern. Menschen neigen dazu, ein präferiertes mentales Modell zu konstruieren und Fehler entstehen dadurch, dass sie keine weiteren Modelle durch Variation dieses Modells berücksichtigen. Diese Befunde basieren aber alle auf Experimenten mit Individuen und wir wissen nur sehr wenig darüber, wie Menschen in Gruppen schlussfolgern. Die Erforschung von Gruppenverhalten hat jedoch beim Problemlösen bereits eine lange Tradition. Ziel des nächsten Abschnitts ist es daher, einen Überblick über Gruppenprozesse

und die Untersuchung von Gruppenentscheidungen zu geben, um anschließend Vorhersagen zum kollaborativen räumlichen Schlussfolgern abzuleiten.

1.4 Gruppenprozesse

Wir haben an dem Beispiel von Sherlock Holmes und Dr. Watson gesehen, dass die Gruppe ein epistemisches Werkzeug sein kann, welches den Individuen beim Denken überlegen ist. Aber nicht nur in der Literatur, sondern auch im Alltag verlassen wir uns oft auf Gruppen. Sei es bei der Entscheidung über eine mögliche Beschäftigung eines neuen Mitarbeiters oder bei einer Krisensitzung zur Klimapolitik. Gruppen gelten als wichtiges Instrument, um komplexe Aufgaben zu bewältigen. Trotz der großen Bedeutung von Gruppen für unsere Gesellschaft gibt es bisher kaum psychologische Untersuchungen zum kollaborativen Schlussfolgern. Deshalb wissen wir nur sehr wenig darüber, welche Prozesse das Denken in Gruppen erklären und wie sich diese von Individuen unterscheiden. Anders als in der Forschung zum Schlussfolgern, gibt es bereits eine große Menge an Forschungen zu Gruppenprozessen in anderen Bereichen der höheren Kognitionen: der *Entscheidungsfindung* und dem *Problemlösen*. Bei diesen Aufgaben werden keine Informationen aus vorhandenem Wissen geschlussfolgert. Deshalb unterliegen sie nicht den Prozessen der Deduktion und ihre Theorien können ganz andere Vorhersagen darüber treffen, wie Individuen oder Gruppen zu einer Lösung gelangen. Das Schlussfolgern unterscheidet sich also grundsätzlich in den kognitiven Prozessen vom Problemlösen und Entscheiden. Trotzdem können wir aus den Theorien und Methoden zu Gruppenprozessen beim Entscheiden und Problemlösen viel lernen. So helfen uns auf der einen Seite die Effekte aus der Entscheidungsfindung, den Einfluss der Gruppe auf das Individuum besser zu verstehen und wir können daraus optimale Rahmenbedingungen für unsere Experimente zum kollaborativen Schlussfolgern ableiten. Auf der anderen Seite bietet uns die Forschung zum Problemlösen in Gruppen wichtige Erkenntnisse, wie man den Leistungsunterschied zwischen Gruppen und Individuen vorhersagen und vergleichen kann.

Deshalb möchte ich im folgenden Abschnitt die wichtigen Konzepte und Theorien zum Entscheiden und Problemlösen beschreiben. Dabei erkläre ich, worauf Gruppenleistungen zurückzuführen sind und wie man diese vorhersagen und experimentell prüfen kann. Anschließend stelle ich empirische Befunde zu diesen Grundlagen dar. Dabei stelle ich zunächst eine gut untersuchte Aufgabe – das *Buchstaben-zu-Zahlen Problem* – zum Problemlösen vor. Bei dieser Aufgabe müssen Individuen und Gruppen durch mathematische Ableitungen einen Code lösen. Diese Aufgabe enthält bereits einige deduktive Komponenten

und kann uns zeigen, warum Gruppen bessere Leistungen beim kollaborativen Denken erzielen können. Anschließend gehe ich auf erste psychologische Befunde von Studien ein, die sich bereits mit Aufgaben zum Schlussfolgern auseinandergesetzt haben, sich jedoch auf das kollaborative Denken mit Konditionalen und Syllogismen konzentrierten. Die Befunde und die theoretischen Annahmen zum Problemlösen setze ich anschließend mit dem räumlichen Schlussfolgern in Bezug und leite Hypothesen zum kollaborativen räumlichen Schlussfolgern ab.

1.4.1 Entscheidungsfindung und Urteilen in Gruppen

Bei der Entscheidungsfindung müssen Gruppen Urteile fällen, bei denen oft keine eindeutig richtige Antwort möglich ist (z. B. Jury-Urteile) und die Forscher sind daran interessiert, wie die Gruppe zu einer Entscheidung gelangt (z. B. Tindale, Nadler, Krebel, & Davis, 2004). Das bedeutet also – im Gegensatz zum Schlussfolgern – dass in den meisten Fällen überhaupt keine richtige Antwort definiert werden kann. Somit hat die Gruppe im Vergleich zum Schlussfolgern nicht das Ziel, aus einem bestehenden Wissenssystem korrekte Ableitungen durchzuführen. Sie versucht vielmehr, einen Konsens aus verschiedenen Meinungen zu finden (Kerr & Tindale, 2004). Dabei ist es vor allem wichtig, wie Informationen und Meinungen innerhalb einer Gruppe verteilt sind. Einer der bekanntesten Befunde stammt dabei von Stasser und Titus (1985). Sie manipulierten die Informationen der einzelnen Gruppenmitglieder so, dass es *geteilte* und *ungeteilte* Informationen gab. Geteilte Informationen sind allen Gruppenmitgliedern gleichermaßen bekannt. Ungeteilte Informationen stehen jedoch nur einzelnen Mitgliedern zur Verfügung. Das bedeutet, dass durch die ungeteilten Informationen ein *hidden profile* (verstecktes Profil, Stasser, 1988) entsteht. Um das *hidden profile* aufzulösen und eine optimale Entscheidung treffen zu können (z. B. die Unschuld eines Angeklagten), müssen die Gruppenmitglieder alle ungeteilten Informationen miteinander austauschen. Das heißt, dass sie ihre eigene Repräsentation teilen müssen, damit die Gruppe gemeinsam eine vollständige Repräsentation konstruieren kann. Stasser und Titus (1985) konnten jedoch zeigen, dass dies nicht der Fall ist. Menschen diskutieren überwiegend diejenigen Informationen, die bereits allen Mitgliedern bekannt sind und es gelingt ihnen nicht, das versteckte Profil zu überwinden. Diese Erkenntnis hatte weitreichende Folgen für die Erforschung von Gruppenprozessen (z. B. Stasser & Abele, 2019) und ist auch für die Untersuchung zum kollaborativen räumlichen Schlussfolgern zu berücksichtigen. So muss klar definiert sein, welches der Gruppenmitglieder welche Informationen erhalten hat und ob die Gruppenmitglieder die Möglichkeit haben, die

Informationen miteinander zu kommunizieren. Außerdem gibt es Faktoren, die der Gruppe helfen können, ein *hidden profile* zu überwinden. So ist es hilfreich, wenn einer Gruppe ausreichend Diskussionszeit zur Verfügung steht (Bowman & Wittenbaum, 2012) und wenn man einen Dissens unter den Gruppenmitgliedern erzeugt, um das Diskutieren zu fördern (Schulz-Hardt, Brodbeck, Mojzisch, Kerschreiter, & Frey, 2006).

Dass aber gerade die Erwartung an einen Dissens problematisch sein kann, zeigt das Phänomen, welches von Janis (1972, 1982) als *Gruppendenken* bezeichnet wurde. Beim Gruppendenken neigen Individuen dazu, ihre Meinung einer (erwarteten) Gruppenmeinung anzupassen und können so als Gruppe schlechtere Entscheidungen treffen als wenn die Individuen sich alleine mit der Aufgabe beschäftigt hätten. Dies wird nach Janis vor allem dadurch motiviert, erwarteten Dissens – und somit Konflikte – in der Gruppe zu minimieren. Dadurch, dass jeder seine Meinung der Gruppenmeinung anpasst, wirkt es für die Gruppe so, als wären alle gleicher Meinung. Dieser Effekt führt nach Janis (1982) zu einer Illusion von einer unantastbaren und fehlerfreien Entscheidung, sodass die Gruppe kritische Evidenzen nicht berücksichtigt oder bewusst ignoriert. Dies gilt als einer der treibenden Faktoren dafür, dass im Jahre 1986 bei der Planung des Starts vom Spaceshuttle „Challenger“ kritische Informationen vernachlässigt wurden, was schlussendlich sieben Astronauten das Leben kostete (Janis, 2008; Moorhead, Ference, & Neck, 1991).

Beim Gruppendenken sind die entstehenden Symptome, wie die empfundene Unantastbarkeit der Gruppenentscheidungen, der treibende Faktor, weshalb die Gruppe zu schlechten Entscheidungen gelangt. Einen sehr ähnlichen Effekt beschrieb Asch (1951) in seinem berühmten Experiment zur Einschätzung von Linienlängen in Gruppen. Er konnte zeigen, dass Individuen bereit waren, ihre eigenen Einschätzungen zugunsten einer falschen Gruppenmeinung anzupassen. Das bedeutet, dass ein Gruppenmitglied zwar die korrekte Lösung wusste, aber auf Grund der Majorität an einer falschen Meinung anderer Mitglieder seine Überzeugung revidiert hat. Dieser *Konformitätseffekt* löste eine ganze Welle sozialpsychologischer Untersuchungen aus. Matz und Wood (2005) zeigten, dass gegensätzliche Meinungen in Gruppen kognitive Dissonanz (Festinger, 1957) erzeugen. Dabei versuchen Personen, diese Dissonanz zu lösen, indem sie sich der Gruppe anpassen (Nir, 2004). Dieser Befund gewinnt in der heutigen Zeit durch die sozialen Medien und Aggressionen im Netz umso mehr an Aktualität. So konnten Neubaum, Rösner, Ganster, Hambach und Krämer (2018) nachweisen, dass Personen mehr Bereitschaft zeigten, einen Selbstjustiz-Aufruf bei Facebook zu unterstützen, wenn der Aufruf viele „Likes“ und Kommentare enthielt.

Gruppendenken ist also der Effekt, dass die Gruppenmitglieder vorab ihre Meinung einer von ihnen erwarteten Gruppenmeinung anpassen und andere Perspektiven innerhalb der Gruppe nicht diskutieren. Konformität entsteht, wenn die Majorität der Gruppenmitglieder eine falsche Antwort vorschlägt und die Minorität (z. B. einzelne Personen) ihre ursprünglich korrekte Antwort zugunsten der Gruppe revidieren. Beide Effekte sind deshalb beim experimentellen Untersuchen von Gruppenleistungen wichtige Faktoren, die berücksichtigt werden müssen. Daher ist es notwendig, die Bedingungen so zu schaffen, dass die Effekte minimal gehalten werden. Hierdurch lassen sich Fehler durch gruppenspezifische Effekte minimieren und die Leistung der Gruppe genauer messen. Doch wie wurden Gruppenleistungen bisher untersucht? Können Gruppen bessere Leistungen als Individuen erzielen und worauf sind potentielle Unterschiede zurückzuführen? Diese Fragen wurden ausführlich beim Problemlösen untersucht und bieten für das Erforschen von kollaborativer Deduktion empirisch, aber auch methodisch wertvolle Hinweise.

1.4.2 Problemlösen in Gruppen

Problemlösen klingt im Alltag sehr ähnlich zum Schlussfolgern, ist aber ein gänzlich anderer Prozess. Beim Problemlösen möchte man etwas von einem Ausgangszustand in einen Zielzustand überführen (z. B. aktuellen Temperaturanstieg des Klimas auf 1.5 Grad begrenzen) (Dunbar & Fugelsang, 2006). Dies wird aber durch Barrieren verhindert und ein Problem entsteht. Personen müssen dann einen Weg finden, den Zielzustand zu erreichen. Das heißt, dass das Problemlösen im Vergleich zum Schlussfolgern kein epistemischer Prozess ist. Beim Schlussfolgern werden anhand von vorhandenem Wissen Informationen abgeleitet, weshalb man beim Schlussfolgern auch nicht von korrekten Lösungen, sondern von *validen Konklusionen* spricht. Dagegen spricht man beim Problemlösen von korrekten Lösungen, um die Barriere des Problems zu überwinden. Wie die Barrieren überwunden werden, ist beim Problemlösen nicht eindeutig vorbestimmt und es sind verschiedene Strategien möglich (Öllinger, 2017). Schlussfolgern kann daher – muss aber nicht notwendigerweise – während des Problemlösens vorkommen (Dunbar & Fugelsang, 2006). Trotz dieser Unterschiede haben Problemlösen und Schlussfolgern in Gruppen auf experimenteller Ebene deutliche Gemeinsamkeiten. So möchte man herausfinden, ob Gruppen bessere Leistungen erzielen als Individuen und wie die Gruppe zu einer Lösung kommt. Methodisch bedeutet dies, dass man Individuen oder Gruppen mit einer Aufgabe (Problem oder Schlussfolgerungen) konfrontiert und anhand der korrekten Antworten bzw. validen Konklusionen misst, ob die Gruppe den Individuen überlegen ist.

Die Leistung des *kollaborativen Problemlösens* ist stark von der Art der zu lösenden Aufgabe abhängig. Deshalb entwickelte Steiner (1966, 1972) eine Taxonomie, die es ermöglicht, eine Aufgabe einer Kategorie zuzuordnen und davon theoretische Vorhersagen über die Gruppenleistung abzuleiten. *Additive* Aufgaben zeichnen sich dadurch aus, dass keine Interaktion notwendig ist und die Einzelleistungen lediglich aufsummiert werden (z. B. *Brainstorming*). Das bedeutet, dass die Überlegenheit der Gruppe lediglich durch Aufsummieren individueller Leistungen erklärt wird. Ähnlich sind *komplementäre* Aufgaben, bei denen verschiedene Einzelleistungen in Subaufgaben aufgeteilt und am Ende aufsummiert werden. Die Leistung der Gruppe kann daher unmöglich von einzelnen Personen erreicht werden (z. B. ein Orchester). Unter *kompensatorischen* Aufgaben versteht Steiner die Aggregation von Einzelleistungen (ohne Interaktion) zu einem Gruppenergebnis. Dieser Aufgabentypus wurde besonders durch Surowieckis (2004) Buch „*Die Weisheit der Vielen*“ bekannt und gründet auf einer Beobachtung des berühmten Naturforschers Sir Francis Galton. Dieser zeigte, dass der simple Durchschnitt aus fast 800 unabhängigen Einzelschätzungen viel besser das Gewicht eines Ochsen abbildete als es eine Einzelschätzung konnte. Somit ist nicht einmal eine Interaktion notwendig, sondern der Fehler des Einzelnen wird durch die Menge an Einschätzungen kompensiert. Für die Überlegenheit der Gruppe gegenüber Individuen ist es also bei den drei Aufgabentypen additiv, komplementär und kompensatorisch nicht zwingend notwendig, dass Gruppenmitglieder miteinander interagieren. Dies ist bei den nach Steiner definierten *konjunktiven* und *disjunktiven* Aufgaben anders. Bei *konjunktiven* Aufgaben entspricht die Gruppenleistung immer der des schlechtesten Individuums (z. B. beim Bergsteigen). Eine Gruppe muss also immer dafür Sorge tragen, dass das schwächste Individuum das Problem lösen kann. Bei *disjunktiven* Aufgaben spricht man häufig von sogenannten „Heureka Problemen“. Dies sind Aufgaben, bei denen die Lösung nicht sofort evident ist. Sobald man sie jedoch gefunden hat, ist es ein Leichtes, sie einem anderen Gruppenmitglied zu erklären. Deshalb entspricht die Leistung der Gruppe nach Steiner bei *disjunktiven* Aufgaben der des besten Individuums. Zu *disjunktiven* Aufgaben zählen eine ganze Reihe von Aufgaben wie das Hypothesentesten, logische Rätsel oder mathematische Probleme, bei denen man die Lösungen schrittweise ableiten kann.

Laughlin und Ellis (1986) betonten, dass bei komplexen Aufgaben die Gruppe die individuelle Leistung übertrifft, wenn die Aufgabe und ihre Lösung *demonstrierbar* sind. Dafür muss eine Aufgabe vier Kriterien erfüllen: (1) Die Gruppenmitglieder müssen sich auf ein gemeinsames System für das Problem einigen können. Das bedeutet, dass sie die gleiche Repräsentation des Problems haben und sich einig sind, wie das Problem zu lösen ist. Sie

müssen also eine geteilte Repräsentation des Problems konstruieren. (2) Die Gruppenmitglieder brauchen genügend Informationen, um die Aufgabe lösen zu können. So ist die Relation „A ist links von B“ eindeutig demonstrierbar, „A ist in der Nähe von B“ kann jedoch nicht ohne weitere Informationen eindeutig gelöst werden. (3) Die Mitglieder, welche die Lösung des Problems nicht kennen, müssen genug Verständnis bezüglich des Systems haben, um die Lösung nachvollziehen zu können. (4) Die Personen, welche die Lösung kennen, müssen die Fähigkeit und Motivation haben, sie den anderen Gruppenmitgliedern zu erklären.

Aus den Kriterien zur Demonstrierbarkeit ergibt sich ein interessanter Zusammenhang zu Steiners (1996, 1972) Taxonomie. Nach Laughlin und Ellis (1986) ist eine Aufgabe umso disjunktiver, je besser sie demonstriert werden kann. Das heißt also, dass umso leichter eine Lösung demonstriert werden kann, desto weniger Individuen einer Gruppe müssen die korrekte Antwort finden. Demnach ist eine Aufgabe nicht notwendigerweise konjunktiv oder disjunktiv, sondern wird auf einem „konjunktiv - disjunktiv Kontinuum“ – welches umgekehrt proportional von der Demonstrierbarkeit abhängig ist – bestimmt. Wenn man also ausreichend Informationen über eine Aufgabe hat, kann man anhand der vier Kriterien der Demonstrierbarkeit vorhersagen, ob eine Gruppe die individuelle Leistung übertreffen kann.

Eines der ersten Experimente zu disjunktiven Aufgaben wurde bereits sehr früh von Shaw (1932) untersucht. Sie kritisierte an der damaligen Forschung, dass die Aufgaben, mit denen Gruppenleistungen gemessen wurden, überhaupt keine Gruppenleistungen erfassen konnten. Dies begründete sie damit, dass für die präsentierten Aufgaben keine Interaktion zwischen den Personen notwendig war. Deshalb führte Shaw mehrere Experimente mit disjunktiven Aufgaben wie das „Türme von Hanoi“ Problem durch. Dieses Problem besteht aus drei Stäben und einem Stapel unterschiedlich großer Holzscheiben. Ziel bei diesem Problem ist es, die Holzscheiben von einer Holzstange zu einer anderen Holzstange zu legen. Dieses Ziel wird aber dadurch erschwert, dass niemals eine kleinere Scheibe auf einer größeren liegen darf und die Probanden müssen sich eine Strategie überlegen, in welcher Reihenfolge sie die Scheiben legen. Diese Aufgabe erlaubt eine klare Messung der Leistung (Anzahl notwendiger Schritte) beim Problemlösen. Shaw verglich die Leistung von Individuen mit der von Gruppen bestehend aus vier Personen. Sie zeigte, dass Gruppen deutlich bessere Ergebnisse erzielen als Individuen. Sie erklärte dies damit, dass Personen in der Gruppe die Möglichkeit haben, sich gegenseitig auf Fehler aufmerksam zu machen und diese zu korrigieren (*Fehlerkorrektur*). Diese Studie galt viele Jahre als Basis für die Annahme, dass Gruppen den Individuen bei komplexen Aufgaben durch Fehlerkorrektur

überlegen sind. Erst zwei Jahrzehnte nach Shaws Experiment, als die mathematische Modellierung von psychologischen Prozessen immer mehr in den Fokus der Forschung rückte, prüfte man andere Theorien, um die Gruppenleistungen zu erklären. Diese Theorien bezeichnet man als *soziale Kombinationsmodelle*. Diese Modelle orientieren sich an der Taxonomie nach Steiner und lassen sich mathematisch abbilden. Sie nehmen an, dass Einzelleistungen in der Gruppe auf bestimmte Arten kombiniert werden können. Deshalb ermöglichen sie die Vorhersage von Gruppenleistungen, wenn man die individuelle Leistung kennt.

1.4.2.1 Soziale Kombinationsmodelle

Lorge und Solomon (1955) waren eine der ersten, die den Ansatz der sozialen Kombinationsmodelle verfolgten. Sie gingen davon aus, dass die Leistung der Gruppe anhand der Wahrscheinlichkeit, mit der ein Individuum die Aufgabe lösen kann, modelliert wird. Das bedeutet, dass genau entgegengesetzt der Fehlerkorrektur ein Individuum die Gruppe nicht von Fehlern, sondern von der korrekten Antwort überzeugen muss. Um zu zeigen, dass diese Annahme die überlegene Gruppenleistung besser als die Fehlerkorrektur erklären kann, prüften sie die Ergebnisse aus dem Experiment von Shaw (1932) mit ihrem bekannten *Modell A*:

$$P_G = 1 - (1 - P_I)^k$$

Wobei P_G die Wahrscheinlichkeit ist, dass die Gruppe die korrekte Lösung findet. P_I ist die Wahrscheinlichkeit, dass ein Individuum die Lösung findet. Der Faktor k steht für die Gruppengröße. Wäre die Wahrscheinlichkeit, dass ein Einzelner eine Aufgabe lösen kann bei $P_I = .60$ und die Gruppe besteht aus $k = 2$ Personen, müsste die Gruppe in mehr als 84 % ($84 = 1 - (1 - 0.6)^2$) der Fälle korrekt sein, damit die überlegene Leistung nicht nur auf die höhere Anzahl von Gruppenmitgliedern zurückzuführen ist. Das Modell A konnte die Daten zum Turm von Hanoi Problem von Shaw sehr genau abbilden. Das Modell zeigte also, dass die Überlegenheit der Gruppe nicht zwingend auf einer Fehlerkorrektur basiert, sondern auf der mit jedem zusätzlichen Mitglied steigenden Wahrscheinlichkeit, die richtige Lösung zu finden. Das bedeutet, dass die Art, wie die Antworten einzelner Gruppenmitglieder kombiniert werden, wichtig für die Überlegenheit der Gruppe ist. Wenn eine Person die Gruppe von der korrekten Repräsentation des Problems überzeugen kann, dann sollte die Gruppe den Individuen überlegen sein (Tindale, Smith, Thomas, Filkins, & Sheffey, 1996). Diese Annahme wurde später als das wichtige *truth wins* Prinzip bekannt (Davis, 1973). Eine

ähnliche Annahme ist, dass mindestens zwei Personen einer Gruppe die korrekte Antwort – und somit die gleiche Repräsentation des Problems – konstruieren müssen, um die restlichen Mitglieder zu überzeugen. Dies kann ebenfalls mit dem Modell A (mit $k = 2$ Personen) abgebildet werden und wird als *truth supported wins* (Laughlin, 1980) bezeichnet und kann ab einer Gruppengröße von $k = 3$ geprüft werden.

Die sozialen Kombinationsmodelle lösten eine ganz neue methodische Vorgehensweise aus, um Gruppenprozesse zu beschreiben und es wurden mit der Zeit eine ganze Reihe weiterer Modelle entwickelt (z. B. Thomas & Fink, 1961; Smoke & Zajonc, 1962). Alle diese Modelle haben gemeinsam, dass sie Regeln definieren, wie Gruppen zu Lösungen von Problemen kommen. Diese Regeln können formal abgebildet werden und ermöglichen anhand von (zuvor erhobenen) individuellen Leistungen die Vorhersage von Gruppenleistungen. Davis (1973) verallgemeinerte diese Regeln in seiner einflussreichen *sozialen Entscheidungsschema-Theorie (social decision scheme theory)*, die formal folgende Form hat:

$$P = \pi * D$$

Wobei P wieder die Wahrscheinlichkeit ist, dass die Gruppe die korrekte Lösung findet. π ist ein Verteilungsvektor individueller Lösungswahrscheinlichkeiten. Dieser Vektor wird entweder theoretisch geschätzt oder anhand erhobener Daten von Individuen ermittelt. D stellt das eigentliche Entscheidungsschema dar und wird in Form einer Matrix abgebildet. Sie gibt an, wie die Anzahl der Personen (r), welche die Aufgabe gelöst haben (*solver*) und die, welche die Aufgabe nicht gelöst haben (*nonsolver*) in einer Gruppe verteilt ist und ob sich die Gruppe anhand dieser Verteilung für die korrekte (C) oder inkorrekte (I) Antwort entscheidet. Die Matrix für das Schema *truth wins* bei einer Gruppe von zwei Personen sieht folgendermaßen aus:

r_s, r_n	truth wins	
	C	I
2,0	1	0
1,1	1	0
0,2	0	1

Anmerkung: r_s = Anzahl der Mitglieder, welche die korrekte Lösung kennen; r_n = Anzahl der Mitglieder, welche die korrekte Lösung nicht kennen; C = Wahrscheinlichkeit, dass die Gruppe die korrekte Lösung nennt; I = Wahrscheinlichkeit, dass die Gruppe die falsche Lösung nennt.

Zur Verdeutlichung nehmen wir an, dass ein Experiment durchgeführt wurde, bei dem Individuen in 60 % der Fälle die korrekte Antwort finden konnten. Wenn wir die Leistung für eine Gruppe aus zwei Personen vorhersagen möchten, ergibt sich der Verteilungsvektor $\pi = (0.36, 0.48, 0.16)$. Diese Werte ergeben sich aus den Wahrscheinlichkeiten, dass beide die korrekte Lösung finden (0.6^2), einer die Lösung findet $[(0.6*0.4) + (0.4*0.6)]$ und keiner die Lösung finden kann (0.4^2). Möchte man nun die Gruppenleistung mit dem *truth wins* Schema vorhersagen, ergibt sich folgende Formel:

$$P = (0.36, 0.48, 0.16) * \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} = (0.84, 0.16)$$

Bei einer individuellen Leistung von 60 % korrekter Antworten müsste die Leistung der Gruppe also deutlich von 84 % korrekten Antworten abweichen, damit das *truth wins* Modell verworfen werden kann. Dies entspricht genau dem Modell A von Lorge und Solomon (1955). Der Vorteil an Davis sozialen Entscheidungsschemata ist, dass man die Matrix D beliebig verschiedenen (theoretisch ermittelten) Schemata anpassen und die resultierenden Modelle gegeneinander testen kann. Die Modelle leiten sich zum größten Teil aus Steiners (1966, 1972) Taxonomie und der daraus resultierenden Vorhersage zu Gruppenleistungen ab. *Truth wins* entspricht der Vorhersage für *disjunktive* Aufgaben. *Error wins* bedeutet, dass die Gruppe von der falschen Repräsentation Einzelner überzeugt wird und deshalb eine falsche Antwort nennt. Dies entspricht den *konjunktiven* Aufgaben nach Steiner, bei der die Gruppe nur so gut wie ihr schlechtestes Mitglied sein kann. Ein weiteres Schema ist *majority*. Hier entscheidet sich die Gruppe wie bei einer Abstimmung für die Lösung, die am meisten Stimmen der einzelnen Mitglieder erhält. Das Schema nimmt also lediglich die Verteilung der individuellen Präferenzen an und niemand muss notwendiger Weise von der richtigen Lösung überzeugt werden. Bei einer Gruppe von zwei Personen oder wenn die Präferenzen in gleichgroßen Gruppen gleich verteilt sind, kann man keine Majorität erzeugen und es greift das Subschema *equiprobability* (Davis, 1973). Dieses macht die gleichen Vorhersagen wie *majority*, außer dass bei gleichen Präferenzen der Zufall entscheidet. Ein Überblick über die verschiedenen Matrizen für eine Gruppe mit vier Personen ist in Tabelle 2 gegeben.

Tabelle 2

Vier Entscheidungsschemata nach Davis (1973) für eine Gruppengröße von $k = 4$ Personen

r_s, r_n	truth wins		error wins		majority (equiprobability)	
	C	I	C	I	C	I
4,0	1	0	1	0	1	0
3,1	1	0	0	1	1	0
2,2	1	0	0	1	0.5	0.5
1,3	1	0	0	1	0	1
0,4	0	1	0	1	0	1

Anmerkung: r_s = Anzahl der Mitglieder, welche die korrekte Lösung kennen; r_n = Anzahl der Mitglieder, welche die korrekte Lösung nicht kennen; C = Wahrscheinlichkeit, dass die Gruppe die korrekte Lösung nennt; I = Wahrscheinlichkeit, dass die Gruppe die falsche Lösung nennt.

Die Theorien der Kombinationsmodelle, zum Gruppendenken und der Konformität zeigen, dass die Gruppe ein sehr komplexes System ist und sehr unterschiedlich beschrieben und beeinflusst werden kann. Obwohl alle vorgestellten Theorien einen wichtigen Beitrag zum Verständnis von Gruppen leisten, bieten sie für sich alleine nur Erklärungen für einen kleinen Bereich der Gruppenprozesse. Das bedeutet, dass sie keine umfassende theoretische Grundlage zu Gruppenleistungen in einer Theorie bieten. Eine Herangehensweise, um diese Theorie zu Gruppenprozessen zu bilden, ist die Informationsverarbeitung. Das Modell der Informationsverarbeitung kann als Grundstein der modernen Kognitionspsychologie gesehen werden (Anderson, 2009). Mit der Zeit haben sich viele verschiedene Modelle zur Informationsverarbeitung bei Individuen entwickelt. Sie alle beschreiben jedoch, dass Informationen schrittweise aufgenommen und als abstrakte Repräsentationen weiterverarbeitet werden. Einer der bekanntesten Modelle zur Informationsverarbeitung ist das *Drei-Speicher-Modell* von nach Atkinson und Shiffrin (1968). Dieses Gedächtnismodell beschreibt, wie Informationen zunächst im sensorischen Speicher aufgenommen, dann im Kurzzeitgedächtnis gehalten und ins Langzeitgedächtnis überführt werden. Dieses Modell wurde später von Baddeley und Hitch (1974) und Baddeley (1992) um das berühmte Arbeitsgedächtnismodell erweitert. Durch die ausführliche Forschung der letzten Jahrzehnte wissen wir sehr gut, wie Informationen bei Individuen verarbeitet werden. Indem man diese Erkenntnisse auf die Gruppe überträgt, kann man vorhersagen, ob und vor allem warum Gruppen den Individuen überlegen sein sollten.

1.4.2.2 Informationsverarbeitung in Gruppen

Hinsz, Tindale und Vollrath (1997) postulieren, dass die Gruppe – analog zu Individuen – ein informationsverarbeitendes System darstellt. In ihrer umfassenden Arbeit stellen sie dar, wie man die bereits durchgeführten Forschungen zu Gruppen in eine Theorie der Informationsverarbeitung von Gruppen einordnen kann. Gruppen unterliegen also nach Hinsz et al. (1997) gleichen kognitiven Prozessen wie die Individuen. Sie bilden auch eine integrierte Repräsentation von Problemen, indem sie gemeinsam die Informationen *encodieren*, *speichern* und *abrufen*. Beim Encodieren werden aus den individuellen Repräsentationen die wichtigen Informationen identifiziert und in eine gemeinsame Repräsentation überführt. Dann wird die neu integrierte Repräsentation von den Individuen gespeichert und schließlich für die Antwort abgerufen. Natürlich stehen der Gruppe allein dadurch, dass sie aus mehreren Personen besteht, mehr kognitive Kapazitäten zur Verfügung. So könnte man annehmen, dass die Gruppe allein durch mehr Arbeitsgedächtniskapazität den Individuen überlegen ist. Dieser Zuwachs ist aber nicht rein additiv, da die Gruppe auch Kapazitäten für die Kommunikation und Diskussion benötigt. Dillenbourg und Betrancourt (2006) bezeichnen diese zusätzliche Belastung als *collaboration load*. Es muss also einen weiteren Faktor geben, der die Gruppe nicht nur quantitativ mehr Informationen verarbeiten, sondern auch qualitativ bessere Lösungen finden lässt. Hinsz et al. (1997) postulieren, dass bei der kollaborativen Verarbeitung von Informationen weniger *Variabilität* in den Repräsentationen entstehen. Das heißt, dass durch den Austausch und die Kombination der unterschiedlichen Informationen eine kohärente geteilte Repräsentation gebildet werden kann. Doch welche Faktoren führen zur Reduzierung der Variabilität?

Das Besondere an der Theorie von Hinsz et al. (1997) ist, dass die Autoren die bereits bekannten Theorien zur Entscheidungsfindung, Problemlösen und Kombinationsschemata nutzen, um die Reduzierung der Variabilität als Informationsverarbeitungsprozess zu erklären. Grundsätzlich ist das Ziel einer Gruppe beim Lösen von Aufgaben, bei denen eine eindeutige Lösung gefunden werden kann, sich gemeinsam auf eine korrekte Repräsentation des Problems zu einigen. Die Gruppe versucht also *Konvergenz* (Kolfshoten & Brazier, 2013) zu erreichen. Die Gruppenmitglieder berücksichtigen unterschiedliche Informationen eines Problems und konstruieren deshalb verschiedene individuelle Repräsentationen. Bei den *hidden profiles* (Stasser & Titus, 1985) konnte man sehen, dass Diversität in den Repräsentationen zu schlechten Gruppenentscheidungen führen kann. Beim Problemlösen ist dies anders. Durch die unterschiedlichen Repräsentationen der Individuen kann die Gruppe nach Hinsz et al. (1997) ein tieferes Verständnis des Problems entwickeln und die

Wahrscheinlichkeit steigt, dass eine Person die Lösung der Aufgabe findet. Diese Person sollte in der Lage sein, die anderen Gruppenmitglieder von der korrekten Lösung zu überzeugen, sodass nach dem *truth wins* Kombinationsschema von Davis (1973) die Gruppe zu einer besseren Lösung gelangt. Damit sie sich jedoch auf eine Repräsentation einigen kann, muss die Gruppe zunächst den Dissens überwinden. Ist der Druck nach Konvergenz zu stark, entstehen Effekte wie die Konformität (Asch, 1951) oder Gruppendenken (Janis, 1982) und die Kombination folgt eher dem *majority* Schema (Davis, 1973). Das bedeutet, dass einzelne Gruppenmitglieder zugunsten einer falschen Gruppenantwort ihre Repräsentation zurückhalten. Die Antwort der Gruppe enthält zwar auch weniger Variabilität, basiert aber nicht auf allen individuellen Repräsentationen. Deshalb muss nach Hinsz et al. (1997) ein ausgeglichenes Maß an Bestreben nach Konvergenz und Dissens herrschen, um alle Informationen für die endgültige Repräsentation zu berücksichtigen.

Ein weiterer Faktor, der die Variabilität in Gruppenentscheidungen reduziert, ist die *Akzentuierung* von kognitiven Prozessen (Hinsz et al., 1997). Dabei werden kognitive Verarbeitungsprozesse und Strategien (z. B. das Falsifizieren von Hypothesen), die mehrere Gruppenmitglieder zeigen, verstärkt und von den anderen Gruppenmitgliedern übernommen. Eine Strategie, die nur von wenigen Gruppenmitgliedern gezeigt wird, kann sich hingegen nicht so leicht durchsetzen. Die Herangehensweise an ein Problem wird akzentuiert und die Gruppe trifft zuverlässigere Entscheidungen. Das bedeutet für Hinsz et al. (1997) aber auch, dass die Gruppe – wenn die Gruppenmitglieder eine schlechte Strategie wählen – zuverlässig schlechtere Entscheidungen treffen. Im schlimmsten Fall kann das bedeuten, dass eine Person die Gruppe von einer fehlerhaften Strategie überzeugt und die Gruppe deshalb das *error wins* Schema nach Davis (1973) zeigt.

Hinsz et al. (1997) sehen also in der Gruppe ein informationsverarbeitendes System, das durch die Integration verschiedener individueller Repräsentationen ein tieferes Verständnis des Problems entwickelt. Es wird Konvergenz erzeugt, indem die zunächst unterschiedlichen individuellen Antworten in eine Repräsentation überführt werden. Sie zeigen aber auch, dass ein zu starkes Streben nach Konvergenz die Antworten verschlechtern können. Eine konvergente Repräsentation ist nur dann hilfreich, wenn die Gruppe so viele Informationen aus den individuellen Repräsentationen wie möglich berücksichtigt und sich dabei auf die richtige Strategie einigt.

1.4.3 Psychologische Befunde zu Gruppenprozessen

Man kann an der Taxonomie nach Steiner (1966, 1972) und den unterschiedlichen Entscheidungsschemata von Davis (1973) sehen, dass die Leistung einer Gruppe stark von den Anforderungen des Problems und der Kombination von den Antworten abhängt. Deshalb sollen im folgenden Abschnitt verschiedene Aufgaben präsentiert werden. Diese sollen uns helfen, die Leistung von Gruppen beim kollaborativen räumlichen Schlussfolgern vorherzusagen. Zunächst stelle ich eine Aufgabe vor, bei der man gut messen kann, ob die Gruppe nach der Theorie von Hinsz et al. (1997) mehr Informationen als ein Individuum verarbeitet. Diese Aufgabe stammt aus dem Bereich des Problemlösens, enthält aber bereits große Anteile des deduktiven Denkens. Anschließend stelle ich weitere empirische Befunde zu konkreten Schlussfolgerungsaufgaben vor und zeige welche Entscheidungsschemata diesen Denkaufgaben zu Grunde liegen.

1.4.3.1 Kollaboratives Problemlösen

Das Buchstaben-zu-Zahlen Problem hat sich als besonders geeignet herausgestellt, die Informationsverarbeitung von Gruppen zu untersuchen. Das Problem enthält zehn Buchstaben (A bis J) und jedem dieser Buchstaben ist ein Wert (0 bis 9) zufällig zugewiesen. Dabei behalten die Buchstaben die Funktion von Zahlen. Das heißt, wenn der Buchstabe A den Wert fünf hat, würde „AA“ die Zahl 55 bedeuten. Die Werte der Buchstaben sind den Probanden unbekannt und ihre Aufgabe ist es, wie bei einem Code, jeden einzelnen Wert für jeden Buchstaben herauszufinden. Die Probanden postulieren eine Formel (z. B. „ $A + B = ?$ “) und der Experimentalleiter gibt ihnen die Lösung der Gleichung (z. B. C). Das heißt, dass er ihnen nicht den Wert des Buchstabens nennt, sondern nur den Buchstaben, der sich durch die Gleichung ergibt. Daraufhin stellen die Probanden eine Hypothese zu dem Wert auf, der sich auf die vom Experimentalleiter zuvor genannten Buchstaben bezieht. Der Experimentalleiter gibt der Gruppe anschließend Rückmeldung, ob diese Hypothese zutrifft. Ein Durchgang besteht also jeweils aus dem Postulieren einer Gleichung und dem Aufstellen einer Hypothese. Die Probanden können pro Gleichung beliebig viele Buchstaben, Additionen und Subtraktionen in der Formel verwenden:

Beispiel

Herauszufindender Code: A = 4; B = 2; C = 0; D = 9; E = 1; F = 7; G = 5; H = 3; I = 6; J = 8

1. Versuch: $A + D = EH$ ($4 + 9 = 13$)

Da die Addition von zwei Zahlen (von 0 bis 9) immer nur zu einer Zahl im Bereich von 0 bis 18 führen kann, können die Probanden schlussfolgern, dass E eine 1 sein muss und stellen im zweiten Versuch:

2. Versuch: $E + E = B$ ($1 + 1 = 2$)

Und finden so heraus, dass B = 2 sein muss. Nun können die Probanden in weiteren Versuchen sich schrittweise an die Lösung des Codes wagen. Gemessen wird dabei die Anzahl an Durchgängen, welche die Probanden für die Lösung des gesamten Codes benötigen. Das Maximum an Durchgängen liegt bei zehn, sodass das Experiment bei mehr Versuchen abgebrochen wird. Es gibt unzählige Möglichkeiten, die Aufgabe zu lösen. So könnte ein spitzfindiger Proband direkt bei der ersten Gleichung auf die Idee kommen, alle Buchstaben (A bis J) miteinander zu addieren. Er kann wissen, dass $0 + 1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 = 45$ ist und würde in dem obigen Beispiel direkt den Wert von A und G herausfinden. Oder wenn man wie im Beispiel beim ersten Schritt den Buchstaben mit dem Wert 1 (E) gefunden hat, kann durch Addition von beliebig vielen „E“ die Zahl „123456789“ ($EEEEEEEE + EEEEE + \dots + EE + E = 123456789$) herausgefunden werden. Den Buchstaben für die Null (C) wird dann durch einfaches Schlussfolgern herausgefunden. Somit ist es möglich, den gesamten Code in nur zwei Schritten zu lösen. Was das Buchstaben-zu-Zahlen Problem so interessant für die Forschung macht ist, dass es zusätzlich zum Hypothesentesten, andere vielseitige Denkstrategien ermöglicht. So kann man durch *Deduktion* wie im Beispiel den Wert von eins (E) herausfinden: „Wenn zwei Buchstaben miteinander addiert werden und das Ergebnis zwei Buchstaben enthält, dann muss der erste Buchstabe des Ergebnisses eine eins sein“ ($A + D = EH$ im Beispiel). Außerdem ist es möglich, durch *transitive Relationen* mehr über die Buchstaben herausfinden, ohne dass die eigentlichen Werte bekannt sind (bei $A + B = D$ muss D größer A und B sein). Man kann also durch deduktive Schlussfolgerungen die Anzahl der benötigten Schritte deutlich minimieren. Neben der Anzahl verwendeter Schritte kann man mit der Aufgabe auch prüfen, ob Gruppen

mehr Informationen verarbeiten. So kann man in einer Gleichung mehr Buchstaben verwenden, um gleichzeitig mehr Informationen zu gewinnen. Diese Informationen können – nach der Informationsverarbeitungstheorie von Hinsz et al. (1997) – von unterschiedlichen Gruppenmitgliedern verarbeitet und anschließend in eine integrierte Repräsentation (die Gleichung) überführt werden. Eine Gruppe hat also die Möglichkeit, die individuelle Leistung deutlich zu übertreffen, wenn sie mehr Informationen gleichzeitig nutzen und bessere Strategien verwenden.

Laughlin, Bonner und Miner (2002) prüften genau diese Annahmen. Sie ließen Individuen oder Gruppen bestehend aus vier Personen das Buchstaben-zu-Zahlen Problem lösen. Dabei maßen sie die Anzahl der verwendeten Schritte, die verwendeten Informationen (Buchstaben) in einer Gleichung und die Menge an identifizierten Buchstaben pro Gleichung. Außerdem betrachteten die Autoren die Anzahl von vorgeschlagenen Gleichungen, bei denen die Probanden das Ergebnis auf der rechten Seite bereits kannten und somit deduktive Schlüsse auf die Werte der Buchstaben links von der Gleichung durchführen konnten (z. B. „Wenn..., dann... Regel bei der Addition von zwei Buchstaben). Sie verglichen diese Leistungen jeweils mit den besten, zweitbesten, drittbesten und viertbesten Individuen der Stichprobe.

Die Ergebnisse waren eindeutig. Gruppen waren in jeder gemessenen Kategorie besser als das beste Individuum. Das heißt, dass die Gruppe mehr Informationen verarbeitet, bessere Strategien verwendet und weniger Schritte für die korrekte Lösung des Codes benötigt hat. Die Ergebnisse stimmen also mit den Annahmen über die Gruppe als Informationsverarbeitungssystem von Hinsz et al. (1997) überein. Die Autoren führen die Ergebnisse darauf zurück, dass die Aufgabe alle vier Kriterien der Demonstrierbarkeit nach Laughlin und Ellis (1986) erfüllen. So können die Probanden sich auf ein System einigen, sie haben ausreichend Informationen zur Verfügung und können sich gegenseitig die korrekten Gleichungen zeigen. Deshalb haben sie die Möglichkeit, mehr Informationen zu verarbeiten und bessere Leistungen zu erzielen.

In einer weiteren Untersuchung testeten Laughlin, Zander, Knievel und Tan (2003), ob Gruppen und Individuen davon profitieren, wenn man sie a) dazu auffordert, eine Mindestzahl an Buchstaben pro Gleichung zu verwenden und b) ihnen den Wert eines Buchstabens vorab sagt. So konnten sie testen, ob Gruppen wirklich mehr Informationen verarbeiten oder Individuen lediglich nicht auf die Idee kamen, mehr Buchstaben pro Gleichung zu verwenden. Sie gaben den Probanden fünf verschiedene Instruktionen: (1) ohne zusätzliche Information (2) mindestens drei Buchstaben pro Gleichung verwenden (3) mindestens vier Buchstaben

verwenden (4) den Buchstaben mit dem Wert Eins vorgegeben (5) den Buchstaben mit dem Wert Neun vorgegeben. Wenn man die Eins vorgibt, erlaubt es theoretisch die im Beispiel vorgestellte Strategie, durch Addition von vielen Einsen die Lösung des Codes direkt im ersten Durchgang zu finden, indem man die Zahl 123456789 generiert und den Buchstaben für Null am Ende schlussfolgert. Wenn man die Neun vorgibt, kann man zwar nicht so leicht in einem Durchgang die Lösung finden, aber ebenfalls Strategien entwickeln, welche die Lösung des gesamten Codes in nur vier Durchgängen ermöglicht.

Die Ergebnisse zeigen wieder eine deutliche Überlegenheit der Gruppe. Gruppen benötigten weniger Schritte als die besten Individuen. Dabei konnten sie bei den Instruktionen, mindestens drei oder vier Buchstaben zu verwenden, die zusätzlichen Informationen deutlich besser verarbeiten. Außerdem nutzten Gruppen oft sogar mehr Informationen, als durch die Instruktion vorgegeben wurde. Individuen hielten sich häufiger an die minimale Vorgabe. Ein weiterer Befund ist, dass Gruppen im Allgemeinen, aber insbesondere bei der vorgegebenen Zahl Neun, effektivere Strategien nutzen und somit schneller den Code lösen konnten.

Die Autoren erklären die Ergebnisse wieder dadurch, dass das Problem die vier Bedingungen der Demonstrierbarkeit erfüllt und Gruppen bessere Informationsverarbeitungssysteme als Individuen sind. Sie betonen aber zusätzlich, dass die Ergebnisse ebenfalls durch überlegene logische Schlussfolgerungen der Gruppe erklärt werden können. Da sie aber die Schlussfolgerungen nicht direkt gemessen, sondern bei den Gruppendiskussionen nur beobachtet haben, können sie diese Erklärung nicht empirisch belegen. Um Schlussfolgerungen bei Gruppen direkt zu messen, haben sich andere – konkret für das deduktive Denken konstruierte – Aufgaben etabliert. Diese wurden bisher im kollaborativen Kontext noch nicht so ausführlich untersucht wie die Aufgaben zum Problemlösen. Die Experimente können jedoch gute Hinweise für die Untersuchungen zum kollaborativen räumlichen Schlussfolgern liefern.

1.4.3.2 Kollaborative Deduktion

Eine bereits sehr gut untersuchte Aufgabe zum deduktiven Denken stellt die *Wason Selektion Task* (Wason, 1966) dar. Bei dieser Aufgabe werden den Probanden vier Karten präsentiert. Auf jeder Karte befindet sich entweder ein Buchstabe oder eine Zahl. Den Probanden wird gesagt, dass sie eine konditionale Regel prüfen sollen: „Wenn ein Vokal auf der einen Seite der Karte ist, dann ist eine gerade Zahl auf der anderen Seite“. Dabei müssen die Probanden

entscheiden welche Karten sie umdrehen, um diese Regel zu prüfen. Sie sind dabei angehalten, die minimale Anzahl an Karten zu wählen, die zur korrekten Lösung führt.

Beispiel

Prüfen Sie folgende Regel: „Wenn ein Vokal auf der einen Seite der Karte ist, dann ist eine gerade Zahl auf der anderen Seite“



Man kann die Regel nur logisch korrekt prüfen, indem man die Karte mit dem Vokal (A) und der ungeraden Zahl (7) umdreht. Nur diese Prüfung entspricht den korrekten Schlusschemata der konditionalen Logik (Modus Ponens und Modus Tollens). Die meisten Probanden entschieden sich jedoch dafür, das A und die Vier umzudrehen. Indem man das A umdreht, kann man direkt die konditionale Aussage prüfen (Modus Ponens). Wenn man die Vier umdreht, hätte es jedoch keinen informativen Mehrwert, da ein Konsonant auf der anderen Seite die Regel nicht falsifizieren würde. Nur wenn hinter der ungeraden Zahl Sieben ein Vokal ist, kann die Regel eindeutig geprüft werden (Modus Tollens). Die *Wason Selection Task* wurde in sehr vielen Variationen getestet und es fällt Menschen sehr schwer, die korrekte Lösung zu finden (z. B. Evans, 2017). Deshalb liegt die Frage auf der Hand, ob Gruppen bei dieser Aufgabe den Individuen überlegen sind.

Moshman und Geil (1998) führten ein Experiment durch, bei dem sie die Leistung von Gruppen mit der individuellen Leistung verglichen. Zusätzlich gab es bei ihnen noch eine Bedingung, in der die Probanden die *Wason Selection Task* erst alleine lösten und dann ein weiteres Mal in einer Gruppe diskutierten und sich für eine gemeinsame Antwort entschieden (Individuum/Gruppe). Dies ermöglichte ihnen nicht nur einen direkten Vergleich der Leistungen von Individuen und Gruppen. Sie konnten zudem verschiedene Gruppenkompositionen betrachten. So ist es möglich, Gruppen zu identifizieren, bei denen die Individuen zuvor keine korrekte Lösung finden konnten. Oder aber man konnte Gruppen betrachten, bei denen mindestens eine Person die Lösung zuvor alleine herausgefunden hatte und diese nach dem *truth wins* Prinzip die Gruppe überzeugen konnte.

Die Ergebnisse waren eindeutig. Nur 9 % der Individuen konnten die Aufgabe lösen. Bei den Gruppen waren es hingegen mit 70 % deutlich mehr korrekte Lösungen. Bei der Individuum/Gruppen Bedingung konnten 80 % der Gruppen die korrekte Lösung finden, wobei der Unterschied zu normalen Gruppen nicht signifikant war. Daraus schlossen die

Autoren, dass die Leistung nicht auf einen Lerneffekt zurückzuführen ist. Am interessantesten sind jedoch die Befunde zu den Gruppenkompositionen. Wenn eine Gruppe mindestens ein Individuum mit der korrekten Antwort enthielt, konnte dieses die Gruppe in 70 % der Fälle von der richtigen Antwort überzeugen, was für das *truth wins* Prinzip spricht. Die Autoren konnten aber zeigen, dass alle Gruppen, bei denen nicht ein Individuum zuvor die korrekte Lösung herausfand, als Gruppe die Lösung finden konnte. Daraus schlussfolgerten die Autoren, dass der Gruppenprozess mehr als nur eine Summe individueller Leistungen sein muss.

Moshman und Geil (1998) begründeten die Überlegenheit der Gruppe vor allem durch die Argumentation und Diskussion der möglichen Antworten in der Gruppe. So beobachteten sie, dass die Personen ihre Ideen immer wieder gegenseitig in Frage stellten und somit gemeinsam eine kohärentere Repräsentation des Problems konstruieren konnten. Ihre Annahmen beziehen sich dabei aber vor allem auf die Beschreibung der Diskussionen und bieten nur eine deskriptive Antwort auf die Überlegenheit der Gruppe. Sie merkten an, dass nicht sicher ist, ob durch die Gruppendiskussion lediglich die individuelle kognitive Fähigkeit gesteigert wird oder ob durch die Diskussion neue, bessere epistemische Strategien zum Vorschein kommen. Wenn lediglich die individuelle Leistung durch den Gruppenkontext gesteigert wird, ist die bessere Leistung nicht direkt auf eine Gruppenleistung zurückzuführen und in einer späteren Testung von Individuen sollte kein Leistungszuwachs erkennbar sein. Wenn aber durch die Gruppe andere Denkprozesse ermöglicht werden, könnten diese Prozesse möglicherweise auch auf spätere individuelle Leistungen übertragen werden.

Genau diese Annahme testeten Maciejovsky und Budescu (2007) in ihrem ersten Experiment mit drei verschiedenen Phasen. In der ersten Phase bearbeiteten die Probanden die *Wason Selection Task* alleine. In der zweiten Phase bearbeiteten sie eine andere Version der *Wason Task* in Gruppen bestehend aus vier Personen. Diese Aufgabe war der Aufgabe in Phase 1 inhaltlich aber noch sehr ähnlich. In der dritten Phase lösten die Probanden wieder alleine eine etwas veränderte Version der *Wason Task*. Die Versionen der *Wason Task* in der ersten und dritten Phase unterschieden sich inhaltlich, sodass zum Beispiel anstelle von Buchstaben und Zahlen geometrische Figuren verwendet wurden. Außerdem wurden die Positionen der korrekten Karten verändert. So konnten die Autoren mit den drei Phasen testen, (1) ob Gruppen besser als Individuen sind und (2) ob es einen Transfer von der Gruppe auf das Individuum gibt.

Ihre Ergebnisse zeigen, dass wie bei Moshman und Geil (1998) in der ersten Phase die Individuen in nur 9 % der Fälle die korrekte Antwort finden konnten und dass Gruppen den

Individuen mit 50 % korrekten Antworten deutlich überlegen waren. Sie fanden außerdem, dass sechs Gruppen die Bedingungen für das *truth wins* Prinzip erfüllten. Das bedeutet, dass bei allen sechs Gruppen mindestens ein Individuum vorhanden war, welches zuvor die korrekte Antwort finden konnte. Von den Gruppen, bei denen keine der Individuen zuvor eine korrekte Lösung fand, konnten 30 % trotzdem die korrekte Antwort durch die Gruppendiskussion finden. Somit konnten Maciejovsky und Budescu (2007) die Befunde von Moshman und Geil (1998) überwiegend replizieren. Das interessanteste Ergebnis ist aber, dass es einen Transfer von der Gruppe auf spätere individuelle Leistungen gab. Sie verglichen die Leistung der ersten Phase bei Personen, die keine Lösung für die *Wason Selection Task* finden konnten, mit ihrer Leistung in der dritten Phase und zeigten, dass über 25 % dieser Personen die richtige Antwort finden konnten. Dieses Ergebnis kann nicht ausschließlich über einen Lerneffekt durch die Wiederholung des Problems erklärt werden, da die Versionen der *Wason Selection Task* sich deutlich unterschieden haben. Maciejovsky und Budescu (2007) gehen davon aus, dass die bessere Gruppenleistung und der Transfer auf ein *geteiltes mentales Modell* zurückzuführen sind. Die Gruppe strebt nach Konvergenz. Sie konstruiert ein gemeinsames Modell der *Wason Selection Task* und akzentuiert die Art, wie diese zu lösen ist. Diese Strategien können die Probanden dann für spätere individuelle Aufgaben übernehmen. Diese Annahme konnten Maciejovsky und Budescu (2007) in weiteren Experimenten stärken indem sie zeigten, dass eine hohe Konvergenz der Gruppenentscheidung mit einem stärkeren Transfer auf die dritte Phase einhergeht.

Moshman und Geil (1998) und Maciejovsky und Budescu (2007) konnten in ihren Untersuchungen zeigen, dass die Gruppe bessere Schlussfolgerungen beim deduktiven Denken ziehen kann als Individuen. Bei ihren Experimenten zeigte sich, dass sogar viele Gruppen, bei denen kein Individuum zuvor die Lösung alleine finden konnte, die richtige Antwort fanden. Bei den Gruppen, in denen nur ein Individuum zuvor die Aufgabe alleine löste, konnten alle Gruppen die korrekte Antwort finden. Dies spricht dafür, dass die individuellen Repräsentationen durch das *truth wins* Prinzip in eine Gruppenrepräsentation kombiniert werden. Die Autoren diskutierten zwar verschiedene Entscheidungsschemata, testeten diese aber nicht explizit.

Eine direkte Untersuchung der unterschiedlichen Entscheidungsschemata beim deduktiven Denken führten Tindale, Smith, Dykema-Engblade und Kluwe (2012) durch. Außerdem wollten sie mehr darüber erfahren, welchen Einfluss die Konvergenz der Gruppenrepräsentation auf die Gruppenleistung hat. Sie verglichen die deduktive Leistung

von Gruppen aus fünf Personen mit Individuen und ließen die Teilnehmer wieder erst alleine und dann in einer Gruppe folgende Syllogismen lösen:

	Valide Konklusion	Invalide Konklusion
1. Prämisse	„Alle Demonstranten sind gut für das Land.“	„Einige unmoralische Menschen sind Wissenschaftler“
2. Prämisse	„Einige radikale Linke sind Demonstranten.“	„Alle Wissenschaftler sind Atheisten.“
Konklusion	„Einige radikale Linke sind gut für das Land.“	„Einige Atheisten sind nicht unmoralisch.“

Das Besondere an diesen Syllogismen ist das Verständnis von dem Wort „*einige*“. In der klassischen Prädikatenlogik kann „*einige*“ auch „*alle*“ bedeuten. Im alltäglichen Verständnis ist dies aber anders. Durch die Implikatur nach Grice (1975) nehmen Personen an, dass „*einige*“ automatisch auch „*nicht alle*“ bedeutet (Begg & Harris, 1982). In dem Beispiel mit der validen Konklusion ist es irrelevant, wie man das Wort *einige* versteht. Bei der invaliden Konklusion ist dies anders. Wenn man impliziert, dass das Wort „*einige*“ in der ersten Prämisse ebenfalls „*nicht alle*“ bedeutet, führt dies dazu, dass man annimmt, dass einige unmoralische Menschen *keine* Wissenschaftler sind. Da die zweite Prämisse besagt, dass alle Wissenschaftler Atheisten sind, schlussfolgert man fälschlicherweise, dass die Konklusion valide sein muss.

Tindale et al. (2012) gingen davon aus, dass Gruppen nach Konvergenz streben und deshalb durch das Teilen ihrer Repräsentationen zu einer Antwort gelangen. Das bedeutet, dass Gruppen bei validen Konklusionen häufiger die korrekte Antwort finden als Individuen. Bei invaliden Konklusionen sollte dies anders sein. Da mehr Personen in der Gruppe durch die fälschliche Implikatur von „*einige*“ eine falsche Repräsentation des Problems haben, wird die falsche Antwort bevorzugt. Das bedeutet, dass die präferierten Repräsentationen der Gruppenmitglieder im Sinne der Akzentuierung nach Hinsz et al. (1997) zu weniger Variabilität gebracht werden und eine falsche Antwort begünstigt wird.

Die Ergebnisse entsprechen genau dieser Vorhersage. Bei validen Konklusionen antworteten Gruppen häufiger korrekt als Individuen. Bei invaliden Konklusionen zeigten die Gruppen im Vergleich zu Individuen eine schlechtere Leistung. Da die Autoren in der ersten Phase die individuelle Leistung und somit auch die Präferenzen der Individuen ermittelten, konnten sie die Verteilung der Präferenzen in der Gruppe schätzen und prüfen, wie die

individuellen Antworten kombiniert wurden. Wenn bei einer validen Konklusion die Majorität die falsche Antwort präferierte, konnte sich die Minorität mit der korrekten Antwort durchsetzen. Das heißt, dass die Minorität in der Lage war, die Gruppe von der korrekten Repräsentation der Aufgabe zu überzeugen. Es wirkten die *truth wins* und *truth supported wins* Prinzipien. Wenn die Konklusionen jedoch invalide waren, konnte sich die Minorität mit der korrekten Antwort viel seltener durchsetzen. Dies entspricht dem *majority* Prinzip. Da die meisten Personen in der Gruppe die falsche Repräsentation der Aufgabe teilten und die korrekte Antwort viel schwieriger zu verteidigen war, setzte sich die falsche Antwort durch.

Insgesamt kann man anhand der vorgestellten Befunde sagen, dass Gruppen sehr häufig den Individuen beim Lösen komplexer Denkaufgaben überlegen sind. Sie verarbeiten mehr Informationen und konstruieren damit kohärentere Repräsentationen. Die Ergebnisse zeigen aber auch, dass Gruppen Fehler machen können, wenn zu viele Personen die gleiche falsche Repräsentation des Problems konstruieren. Die Befunde zum deduktiven Denken basieren noch auf wenigen einzelnen Experimenten und im Falle der *Wason Selection Task* ist nach wie vor nicht eindeutig geklärt, wie Individuen diese Probleme lösen (Ragni, Kola, & Johnson-Laird, 2017). Dies ist beim räumlichen Schlussfolgern anders. Deshalb möchte ich im nächsten Kapitel als erstes die Gruppe als Untersuchungsgegenstand meiner Arbeit definieren und dann die Erkenntnisse zum kollaborativen Problemlösen und Schlussfolgern mit den Befunden zum individuellen räumlichen Schlussfolgern kombinieren und meine Hypothesen zum *kollaborativen räumlichen Schlussfolgern und Überzeugungsrevision* ableiten.

Kapitel 2: Integration der Theorien und Hypothesen

Kapitel 1 konnte zeigen, dass wir bereits sehr viel darüber wissen, wie Gruppen komplexe Aufgaben lösen und wie sie zu einer gemeinsamen Antwort kommen. Dabei können Gruppen auf der einen Seite Fehler begehen, wenn sie nicht in der Lage sind, alle Informationen miteinander zu teilen und deshalb ein *hidden profile* nicht überwinden können. Auch gruppenspezifische Effekte wie die Konformität und das Gruppendenken können dazu führen, dass die Gruppe fatale Fehlentscheidungen fällen kann. Besonders das Gruppendenken gilt als Ursache für einige der schrecklichsten Unglücke der Menschheitsgeschichte, wie die Explosion des Chernobyl Kernreaktors (Reason, 1987) oder die Challenger Explosion (siehe Kapitel 1.4.1: Gruppenprozesse; Janis, 2008; Moorhead et al., 1991). Auf der anderen Seite zeigen die Forschungen zum kollaborativen Problemlösen, dass Gruppen Informationen sehr oft viel besser verarbeiten können als Individuen. Gruppen sind in der Lage, beim mathematischen „Buchstaben-zu-Zahlen Problem“ mehr Informationen zu berücksichtigen und die individuellen Repräsentationen eines Problems in einer gemeinsamen Repräsentation zu kombinieren (Laughlin et al., 2002). Die Entscheidungsschemata zeigen, dass Gruppenmitglieder sich gegenseitig von korrekten Repräsentationen der Probleme überzeugen können. Außerdem kann man durch die Schemata sehen, dass häufig nicht einmal eine Person in der Gruppe alleine die korrekte Lösung eines Problems benötigte, damit die Gruppe das Individuum übertreffen kann. Vielmehr können sich die Gruppenmitglieder gegenseitig durch Diskussion von den verschiedenen Perspektiven eines Problems überzeugen und können so folgerichtige Konklusionen ableiten (Moshman & Geil, 1998). Alle diese positiven, aber auch negativen Eigenschaften einer Gruppe werden maßgeblich von der Definition über die Art der Interaktion und die Gruppengröße bestimmt. Deshalb ist es sehr wichtig für meine Arbeit, die Gruppe als Untersuchungsgegenstand zunächst eindeutig zu definieren. Aus dieser Definition ergeben sich theoretische und methodische Implikationen über negative Effekte wie die Konformität, aber auch über die Entscheidungsschemata. Diese Implikationen geben uns Hinweise für die Leistung beim räumlichen Schlussfolgern in Gruppen.

Die Definition der „Gruppe“ kann je nach Forschungsschwerpunkt stark variieren und zu anderen Vorhersagen führen (Witte, 2014). Für meine Arbeit verwende ich deshalb die psychologische Definition der Gruppe von Shaw (1971). Für Shaw besteht die Gruppe aus zwei oder mehr Personen, welche sich gegenseitig durch Interaktion beeinflussen. Dabei ist es zwingend notwendig, dass die Personen interagieren. So kann zum Beispiel eine Person A auf einer Bank an der Nordsee sitzen und den Sonnenuntergang genießen. Wenn sich eine Person

B dazu setzt und das gleiche Phänomen beobachtet, ist dies nach Shaw noch keine hinreichende Bedingung für eine Gruppe. Erst wenn A und B sich beispielsweise durch ein Gespräch gegenseitig beeinflussen – also interagieren – erfüllen sie die hinreichende Bedingung für eine Gruppe. Das Zitat von Dr. Watson in meiner Einleitung stellt sehr gut dar, wie durch die Interaktion zwischen ihm und Sherlock Holmes die herausragenden Denkleistungen entstehen. Dabei ist ein Paar aus zwei Personen wie dieses literarische Duo die kleinste Form der Gruppe und wird als *Dyade* bezeichnet. Dyaden bieten für die experimentelle Untersuchung des Schlussfolgerns spezielle Eigenschaften. Das Besondere an Dyaden ist, dass sie negative Effekte wie den *Ringelmann Effekt* (auch: *soziales Faulenzen*; Moede, 1927; Ingham, Levinger, Graves, & Peckham, 1974) reduzieren. Dieser Effekt besagt, dass die einzelnen Gruppenmitglieder weniger Aufwand für die Lösung eines Problems anwenden, weil sie annehmen, dass die anderen Gruppenmitglieder diese Leistungsreduzierung kompensieren. Dieser Effekt wird stärker, je größer die Gruppe ist (Latané, Williams, & Harkins, 1979) und somit ist es sinnvoll, besonders bei dem neuen Forschungsgebiet des kollaborativen räumlichen Schlussfolgerns die Gruppengröße auf ein Minimum zu reduzieren. Für Witte (2014) kann die Dyade die Definition der Gruppe nicht erfüllen, da in ihr keine Koalition gebildet und somit kein Anpassungsdruck entstehen kann. Deshalb muss nach Witte eine Gruppe aus mindestens drei Personen bestehen. Für die experimentelle Kontrolle ist es jedoch wünschenswert, den Anpassungsdruck zu minimieren. So kann die Leistung beim Schlussfolgern deutlich weniger durch gruppendynamische Effekte, wie zum Beispiel die Konformität (Asch, 1951, 1956), verzerrt werden. Weil in einer Dyade keine Majorität Einfluss auf den Einzelnen nehmen kann, haben Dyaden auch besondere Eigenschaften für die Kombination der Antworten und die Vorhersage der Gruppenleistung kann besser spezifiziert werden. Die sozialen Entscheidungsschemata von Davis (1973) reduzieren sich bei einer Dyade auf drei mögliche Kombinationen. (1) Eine Person überzeugt die andere von der korrekten Antwort (*truth wins*), (2) beide können sich nicht einigen und wählen zufällig die Antwort von einer der beiden Personen (*equiprobability*) oder (3) eine Person überzeugt die andere von der falschen Antwort (*error wins*). Die Dyade reduziert also gruppendynamische Effekte sowie die möglichen Antwortkombinationen. Deshalb bietet die Dyade optimale Eigenschaften, um die Gruppenleistung beim Schlussfolgern in einem möglichst kontrollierten experimentellen Rahmen zu untersuchen. Aus diesem Grund habe ich mich für alle Experimente zum Denken in Gruppen auf den Vergleich zwischen Individuen und Dyaden entschieden.

Mit Hilfe der Gruppendifinition ist es nun möglich, die Erkenntnisse zum individuellen räumlichen Schlussfolgern mit den Theorien zum kollaborativen Problemlösen zu verbinden. Das menschliche Denken folgt den drei Phasen der Konstruktion, Inspektion und Variation mentaler Modelle (Johnson-Laird & Byrne, 1991; Knauff, et al., 1998b). Wir wissen von den Untersuchungen mit Individuen, dass diese zunächst eine integrierte Repräsentation – ein mentales Modell – der Prämissen konstruieren und diese dann nach Informationen inspizieren, die nicht explizit in den Prämissen genannt wurden. Falls die Prämissen weitere Interpretationen zulassen, wird zunächst eine mentale Annotation (Rauh, 2000) über die weiteren Möglichkeiten im Arbeitsgedächtnis gehalten und dann das Modell variiert (Knauff et al., 1998b). Wenn aber neue, inkonsistente Informationen gegen das präferierte Modell sprechen, ist eine besondere Form der Variation, die Revision, notwendig (Knauff et al., 2013). Eine Aufgabe beim Denken ist dabei umso schwieriger, desto mehr Informationen und somit Modelle berücksichtigt werden müssen (Byrne & Johnson-Laird, 1989; Carreiras & Santamaria, 1997; Roberts, 2000, Knauff, 2013). Die Informationsverarbeitung bei Gruppen ist dabei ganz ähnlich zu den Phasen der mentalen Modelle. Wir haben beim kollaborativen Problemlösen und Denken in Gruppen gesehen, dass die Gruppe ebenfalls eine integrierte Repräsentation der Probleme konstruiert. Dabei werden die individuellen Repräsentationen in eine Gruppenrepräsentation – ein geteiltes mentales Modell (Maciejovsky und Budescu, 2007; Hinsz et al., 1997) – überführt. Die Gruppe versucht auf der einen Seite möglichst viele Informationen der individuellen Repräsentation zu übernehmen. Auf der anderen Seite hat die Gruppe als Ziel die Konvergenz (Kolfshoten & Brazier, 2013). Das bedeutet, dass die Gruppe die Variabilität reduzieren möchte (Hinsz et al., 1997) und einzelne Überzeugungen gegebenenfalls zu Gunsten der Gruppenrepräsentation entweder variiert oder sogar revidiert werden. Die Gruppe scheint also die Informationen sehr ähnlich zum individuellen Denken zu verarbeiten. Doch bedeutet dies, dass die Gruppe auch die gleichen Schwierigkeiten beim Denken hat? Sind mehr Informationen genauso schwierig zu verarbeiten und zeigt die Gruppe deshalb auch die gleichen Effekte wie Individuen beim Denken?

Wir wissen von den Experimenten mit Individuen, dass die Phasen der Konstruktion, Variation und Revision beim räumlichen Schlussfolgern durch unterschiedliche Effekte gezeigt werden können. Die Phase der Konstruktion konnte am besten mit dem Kontinuitätseffekt getestet werden. Der Unbestimmtheitseffekt ist am besten geeignet, um die Variation mentaler Modelle zu prüfen und für die Phase der Revision konnte das LO-Prinzip in vielen Experimenten nachgewiesen werden. Wenn die Informationsverarbeitung der

Gruppe analog zu der Verarbeitung beim Denken von Individuen ist, dann sollten auch Gruppen diese Effekte beim kollaborativen Denken zeigen. Diese Annahme habe ich in *drei experimentellen Teilen* getestet. Dabei habe ich die Prozesse der Konstruktion und Variation bei Dyaden und Individuen verglichen. Die Phase der Revision habe ich ausschließlich an Einzelpersonen getestet. Auch wenn die Dyade als kleinste Form der Gruppe gruppenspezifische Einflüsse wie die Konformität minimieren, ist es sehr schwierig den Diskussionsprozess zu kontrollieren. Deshalb habe ich einen fiktiven Gruppenkontext gewählt und untersucht, wie soziale Faktoren die Überzeugungsrevision bei Individuen beeinflussen. In den nächsten drei Abschnitten werde ich deshalb zeigen, welche Hypothesen sich aus der Kombination der Theorien zum räumlichen Schlussfolgern und Gruppenprozessen für die *kollaborative Konstruktion*, *kollaborative Variation* und *soziale Revision* mentaler Modelle ableiten. Die Hypothesen sind dabei so strukturiert, dass sie in jedem Teil für sich alleine stehen. Das bedeutet, dass jeder experimentelle Teil mit einer eigenen Hypothese 1 beginnt. Somit kann jeder experimentelle Teil als eigenständiges Kapitel gelesen und verstanden werden.

2.1 Kollaborative Konstruktion mentaler Modelle

Für den *experimentellen Teil 1: Kollaborative Konstruktion* habe ich den Kontinuitätseffekt verwendet. Der Kontinuitätseffekt entsteht, wenn die Prämissen in unterschiedlich aufeinander aufbauenden (kontinuierlichen, semi-kontinuierlichen oder diskontinuierlichen) Reihenfolgen präsentiert werden (Knauff et al., 1998a). Kontinuierliche und semi-kontinuierliche Probleme können mit jeder weiteren Prämisse schrittweise konstruiert werden. Diskontinuierlich präsentierte Prämissen hingegen bauen nicht aufeinander auf und ein zusammenhängendes Modell ist erst mit der letzten Prämisse möglich. Bei Experimenten mit Individuen zeigte sich, dass diskontinuierliche Aufgaben am schwierigsten zu verarbeiten sind. Das zeigt sich insbesondere in einer deutlich längeren Verarbeitungszeit der dritten Prämisse und außerdem in mehr fehlerhaft konstruierten Modellen (Ehrlich & Johnson-Laird, 1982; Knauff, et al., 1998a). Gruppen konstruieren ihre geteilten mentalen Modelle, indem sie die individuellen Repräsentationen in eine kohärente Repräsentation integrieren (Hinsz et al., 1997). Wenn dieses geteilte mentale Modell von der individuellen Modellkonstruktion abhängig ist, dann sollte die Dyade auch dem Kontinuitätseffekt unterliegen. Meine erste Hypothese lautet deshalb, dass *Dyaden beim gemeinsamen Konstruieren eines geteilten mentalen Modells den Kontinuitätseffekt zeigen*. Das geteilte mentale Modell einer Dyade basiert aber nicht nur auf der Repräsentation einer einzelnen Person. Vielmehr werden die

Informationen von zwei Personen in ein Modell integriert. Das bedeutet, dass die Dyade mehr Informationen verarbeiten kann und zusätzlich weniger Variabilität in ihrem geteilten mentalen Modell erzeugt. Deshalb lautet meine zweite Hypothese, dass *Dyaden mehr korrekte Modelle konstruieren als Individuen*.

Ein anderer wichtiger Befund zum Kontinuitätseffekt kommt von Nejasmic et al. (2015). Sie zeigten, dass Individuen selbst die diskontinuierlichen Probleme von Anfang an als ein zusammenhängendes Modell konstruieren. Das bedeutet, dass die ersten beiden Prämissen auch ohne eine Information über das gesamte Modell in ein vorläufiges Modell integriert werden. Dieses vorläufige Modell enthält eine Annotation darüber, dass die Relation von allen Termen nicht eindeutig bestimmt werden kann. Erst wenn die Probanden die Informationen der dritten Prämisse erhalten, können sie das vorläufige Modell variieren, um das endgültige integrierte mentale Modell zu konstruieren. Nach der mentalen Modelltheorie geht die verbale Information der Prämissen bei der Integration in ein Modell verloren (Knauff, 2013). Das bedeutet, dass selbst in dem vorläufig konstruierten Modell der genaue Wortlaut der Prämissen nicht enthalten ist. Beim Denken in Gruppen ist es notwendig, die individuellen mentalen Modelle untereinander zu teilen. Deshalb ist eine Dyade darauf angewiesen, dass eine Person ihrem Partner so viele und detaillierte Informationen wie möglich mitteilt. Wenn man aber die Informationen der Prämissen in ein vorläufiges Modell integriert und die verbale Information verloren geht, wie teilen Mitglieder einer Dyade dann dieses Modell mit ihrem Partner? Wenn eine Person alleine denkt, findet der Konstruktionsprozess ausschließlich mental statt. Ein Individuum hat nicht das Ziel, seine Repräsentation so detailliert wie möglich mit einem Partner zu teilen. Das bedeutet, – wenn man ein Individuum bittet, sein vorläufig konstruiertes Modell laut aufzusagen – dass es irrelevant ist, ob ein Individuum dabei die Annotation über die Diskontinuität der Prämissen berücksichtigt. In diesem Fall sollte sie ein vorläufiges Modell lediglich als Abfolge von Objekten (z. B. Apfel Birne Mango Kiwi) verbalisieren. Bei der Kollaboration in einer Dyade sollte dies anders sein. Ausgehend von der Theorie der Informationsverarbeitung von Gruppen nach Hinsz et al. (1997) und den Befunden von Maciejovsky und Budescu (2007) nehme ich an, dass Personen in einer Dyade ihrem Partner so viel wie möglich Informationen mitteilen möchten um ein möglichst kohärentes geteiltes mentales Modell zu konstruieren. Wenn eine Person also ihrem Partner das vorläufige Modell mitteilen soll, berücksichtigt sie die Annotation über die Diskontinuität und verbalisiert diese, indem sie beispielsweise „Apfel Birne und Mango Kiwi“ sagt. Somit macht sie die Annotation explizit und zeigt, dass nur die Relationen von Apfel zu der Birne sowie von Mango zur Kiwi eindeutig sind. Deshalb lautet

meine dritte Hypothese, dass *Dyaden bei diskontinuierlicher Prämissenreihenfolge häufiger die Annotation über die Diskontinuität verbalisieren als Individuen*.

Zusammengefasst lauten also die Hypothesen für den experimentellen Teil 1: Kollaborative Konstruktion mentaler Modelle: (1) Dyaden zeigen den Kontinuitätseffekt, (2) konstruieren jedoch mehr korrekte Modelle als Individuen und (3) Mitglieder einer Dyade verbalisieren häufiger die Annotation der Diskontinuität als Individuen. Diese Hypothesen habe ich in zwei Experimenten getestet. *Experiment 1* testete die erste und die zweite Hypothese. Deshalb habe ich die Verarbeitungszeit von Dyaden untersucht und außerdem die konstruierten Modelle von Dyaden mit denen von Individuen verglichen. Im *Experiment 2* habe ich die zweite und dritte Hypothese getestet und verglichen, ob Personen in einer Dyade die vorläufig konstruierten Modelle anders verbalisieren als Individuen.

2.2 Kollaborative Variation mentaler Modelle

Wenn ein Problem beim Denken mehr als eine Interpretation der Informationen zulässt, ist die Variation mentaler Modelle die wichtigste Phase, um valide Konklusionen zu ziehen. Je mehr mentale Modelle für ein Problem berücksichtigt werden müssen, desto mehr Fehler entstehen (Byrne & Johnson-Laird, 1989). Dieser Unbestimmtheitseffekt ist einer der wichtigsten Befunde, um das räumliche Schlussfolgern zu untersuchen. Deshalb liegt der Schwerpunkt meiner Arbeit auf dem Unbestimmtheitseffekt und ist Gegenstand meines *experimentellen Teil 2: Kollaborative Variation*. Wir wissen von den Experimenten mit Individuen, dass sie – auch wenn indetermierte Probleme mehr als eine mögliche Interpretation zulassen – nur ein präferiertes mentales Modell konstruieren (Jahn et al., 2007; Knauff et al., 1995). Anstatt mehrere Modelle zu konstruieren, halten Individuen eine Annotation über die Möglichkeit weiterer Modelle im Arbeitsgedächtnis (Rauh, 2000; Vandierendonck et al., 2004) und können so die kognitive Belastung reduzieren. Problematisch wird dies aber, wenn die Personen diese Annotation nicht weiter berücksichtigen oder vergessen und somit das präferierte mentale Modell nicht weiter variieren (Rauh et al., 2005). Doch was geschieht, wenn Dyaden gemeinsam Denkaufgaben bearbeiten, bei denen mehrere Modelle möglich sind? Auf der einen Seite wissen wir von den Experimenten zum kollaborativen Problemlösen, dass Gruppen mehr Informationen berücksichtigen und verarbeiten können (Laughlin et al., 2002, 2003). Auf der anderen Seite können Gruppen aber durch das Konvergenzbestreben (Kolfshoten & Brazier, 2013) und die Akzentuierung ihrer kognitiven Prozesse (Hinsz et al., 1997) dazu neigen, sich nur auf eine (falsche) von den Individuen präferierte Repräsentation zu einigen (Tindale et al., 2012). Wir haben aber auch gesehen,

dass Gruppen selbst dann, wenn keine ihrer Individuen die korrekte Repräsentation eines Problems bilden konnte, durch die Diskussion eine valide Konklusion ziehen können (Moshman & Geil, 1998; Maciejovsky & Budescu, 2007). Das heißt, dass wir nur anhand der Informationsverarbeitung nicht eindeutig vorhersagen können, ob die Dyade den Individuen beim Unbestimmtheitseffekt überlegen ist.

Wir haben aber im Bereich des Problemlösens gesehen, dass wir durch die Eigenschaften der Aufgabe viel darüber herausfinden können, ob die Gruppe eine bessere Leistung als Individuen erzielen kann. Die Probleme zum Unbestimmtheitseffekt (ein-Modell, zwei-Modelle determiniert und zwei-Modelle indeterminiert) sind komplexe Denkaufgaben und sind nach Steiner (1966, 1972) der *disjunktiven* Kategorie zuzuordnen. Die disjunktive Kategorie sagt grundsätzlich eine Überlegenheit der Gruppe vorher. Dies ist aber nur dann der Fall, wenn sie die vier Kriterien der Demonstrierbarkeit erfüllt (Laughlin & Ellis, 1986). Ausgehend von der Theorie mentaler Modelle von Johnson-Laird und Byrne (1991) und dessen Erweiterung von Knauff (2013) beim individuellen Schlussfolgern lässt sich ableiten, dass (1) die Individuen in der Dyade ein einheitliches repräsentatives System – die mentalen Modelle – verwenden und ihre Repräsentation des Problems miteinander teilen können. (2) Der Dyade stehen durch die Prämissen ausreichend Informationen zur Verfügung, um die Aufgabe zu lösen. (3) Ein Mitglied, welches ein falsches Modell konstruiert, kann durch die Unterstützung des Partners seine Überzeugung revidieren und die Lösung nachvollziehen. (4) Das Mitglied, welches die korrekten Modelle konstruiert hat, ist in der Lage, dem Partner dieses Modell zu erklären oder zu zeigen. Die Probleme beim Unbestimmtheitseffekt sind also eindeutig demonstrierbar und sollten deshalb dazu führen, dass Dyaden mehr valide Konklusionen beim räumlichen Schlussfolgern ziehen können als Individuen. Deshalb nehme ich in meiner ersten Hypothese an, dass *Dyaden mehr valide Konklusionen ziehen als Individuen*. Da sie mehr Informationen verarbeiten können, sollten sie außerdem bei Problemen, die mehr als ein Modell zulassen, häufiger die Annotation weiterer Interpretationen für ihre Antwort berücksichtigen können. Das bedeutet, dass die Dyade besonders dann profitiert, wenn die Probleme die Berücksichtigung von mehreren Modellen erfordern und somit schwieriger sind. Deshalb lautet meine zweite Hypothese, dass *Dyaden besonders bei indeterminierten Problemen mehr valide Konklusionen ziehen als Individuen*.

Um mehr darüber zu erfahren, wie die Gruppe gemeinsam bei indeterminierten Problemen zu einer Konklusion gelangt, können wir die Kombinationsprozesse der Gruppe genauer untersuchen. Wir haben bereits gesehen, dass die Dyade die möglichen Kombinationsprozesse auf die Entscheidungsschemata *truth wins*, *equiprobability* und *error*

wins nach Davis (1973) reduziert. Zum einen macht jedes dieser Schemata Annahmen darüber, ob mindestens eine Person eine korrekte (oder fehlerhafte) Repräsentation der Aufgabe konstruieren kann und die Gruppe davon überzeugt. Zum anderen wissen wir sehr viel darüber, welche (fehlerhaften) Repräsentationen Individuen beim räumlichen Schlussfolgern konstruieren. Wenn die Gruppenmitglieder sich gegenseitig von einem präferierten geteilten mentalen Modell überzeugen und deshalb ebenfalls die Annotation über mehrere Interpretationen der Prämissen ignorieren, dann sollte die Gruppenleistung am besten durch das *error wins* Schema vorhergesagt werden. Wenn sie sich nicht auf ein geteiltes mentales Modell einigen können, sollte das *equiprobability* Schema die Leistung am besten erklären. Wenn sie jedoch in der Lage sind, sich von der korrekten Repräsentation, die ebenfalls die Annotation über weitere alternative Modelle enthält, zu überzeugen, dann sollte die Gruppenleistung auf das *truth wins* Schema zurückzuführen sein. Wir haben in den Untersuchungen zum deduktiven Denken in Gruppen gesehen, dass das *truth wins* Schema die Denkleistungen von Gruppen am besten erklären konnte (Maciejovsky & Budescu, 2007; Moshman & Geil, 1998; Tindale et al., 2012). Deshalb nehme ich an, dass auch beim räumlichen Schlussfolgern sich die Mitglieder der Dyade gegenseitig von dem korrekten mentalen Modell der Probleme überzeugen können und meine dritte Hypothese lautet, dass *die Leistung der Dyade bei indeterminierten Problemen am besten durch das truth wins Schema vorhergesagt wird.*

Zusammenfassend lauten die Hypothesen des experimentellen Teil 2: Kollaborative Variation mentaler Modelle: (1) Dyaden ziehen mehr valide Konklusionen als Individuen und (2) profitieren besonders dann, wenn die Probleme indeterminiert sind. (3) Die Leistung der Dyade wird bei indeterminierten Problemen am besten durch das *truth wins* Schema vorhergesagt. Diese Hypothesen habe ich in drei Experimenten getestet. Alle Experimente vergleichen die Leistung von Individuen mit der Leistung von Dyaden bei ein-Modell, zwei-Modelle determinierten und zwei-Modelle indeterminierten Problemen. Um dabei die einzelnen Vorhersagen zur Integration der mentalen Modelle genauer testen zu können, habe ich für jedes Experiment leichte methodische Variationen vorgenommen. In *Experiment 3* haben die Probanden zunächst individuell ein mentales Modell der Probleme konstruiert und sollten sich dann gemeinsam auf eine korrekte Antwort einigen. Damit sie ihre individuellen Modelle miteinander teilen und diskutieren konnten, standen ihnen reale Objekte zur Verfügung.

In *Experiment 4* konnten die Personen in der Dyade ihre individuellen Repräsentationen nicht durch reale Objekte zeigen. Das geteilte mentale Modell musste also ausschließlich durch

verbale Beschreibungen und Diskussion konstruiert werden. In *Experiment 5* verwendete ich mehrere Phasen, um die Leistung der Dyade mit der der Individuen zu vergleichen. Das heißt, dass die Probanden zunächst alleine und dann in der Dyade (oder umgekehrt) gearbeitet haben. Diese methodische Variation ermöglicht es wie bei den Experimenten von Maciejovsky und Budescu (2007) mehr darüber zu erfahren, ob durch das geteilte mentale Modell in der Dyade kognitive Prozesse entstehen, die auf das individuelle Denken übertragen werden können.

2.3 Soziale Revision mentaler Modelle

Wenn Dyaden ihre individuellen mentalen Modelle beim kollaborativen Schlussfolgern diskutieren, kommt es unweigerlich dazu, dass sich die Gruppenmitglieder gegenseitig mit neuen, inkonsistenten Informationen konfrontieren. Deshalb müssen sie ihre eigene Überzeugung gelegentlich aufgeben. Diese räumliche Überzeugungsrevision ist Inhalt meines *experimentellen Teil 3: Soziale Revision mentaler Modelle*. Die räumliche Überzeugungsrevision ist ein sehr gut untersuchtes Gebiet. Dabei stellte sich heraus, dass die Menschen sehr häufig die funktionelle Asymmetrie der Terme des Faktes – das LO-Prinzip – nutzen (Bucher et al., 2011). Einerseits ist dieses Prinzip so stark, dass es nur sehr schwer beeinflusst werden kann (Bucher et al., 2013; Bucher & Nejasmic, 2012; Knauff et al., 2013; Krumnack et al., 2011). Andererseits haben wir aber gesehen, dass gegensätzliche Meinungen in einer Gruppe starke kognitive Dissonanz erzeugen (Matz & Wood, 2005). Deshalb passen sich Menschen häufig den Gruppenmeinungen an. Diese Konformität kann so stark sein, dass man eine korrekte Meinung zu Gunsten einer falschen Gruppenmeinung aufgibt (Asch, 1951, 1956). Deshalb ist anzunehmen, dass ein starker Konformitätsdruck dazu führen kann, dass Personen das LO-Prinzip zu Gunsten einer anderen Gruppenmeinung aufgeben. Wenn die Gruppe aber ebenfalls das LO-Prinzip bevorzugt, sollte es noch häufiger genutzt werden. Meine erste Hypothese lautet deshalb, dass *die Konformität einen Einfluss auf die Revisionsstrategie beim Revidieren räumlicher mentaler Modelle hat*.

Im Bereich der Überzeugungsrevision mit konditionalen Aussagen zeigten Wolf et al. (2012), dass die Vertrauenswürdigkeit der Quelle bestimmen kann, welche Aussage Personen verwerfen. Entgegen der Vorhersage des *mismatch principle* (Johnson-Laird et al., 2004) entschieden sich Personen die Aussagen zu verwerfen, die von einer wenig vertrauenswürdigen Quelle geäußert wurden. Dieses Quellenprinzip (Wolf et al., 2012) ist also stärker, als das einfache Abgleichen der mentalen Modelle, die mit dem Fakt inkonsistent sind. Wenn Personen bei der räumlichen Überzeugungsrevision das LO-Prinzip verwenden,

geben sie automatisch eine der beiden ursprünglichen Prämissen auf. Wenn diese Prämisse von einer vertrauenswürdigen Quelle stammt, dann sollten die Personen nicht bereit sein, das LO-Prinzip zu verwenden. Hingegen kann das LO-Prinzip verstärkt werden, wenn die Quelle der nach dem LO-Prinzip zu verwerfenden Prämisse nicht sehr vertrauenswürdig ist. Deshalb lautet meine zweite Hypothese, dass *das Quellenprinzip die Revisionsstrategie beim Revidieren räumlicher mentaler Modelle beeinflusst*.

Zusammenfassend lauten also die Hypothesen zum experimentellen Teil 3: Soziale Revision mentaler Modelle: (1) die Konformität beeinflusst das LO-Prinzip und (2) das Quellenprinzip beeinflusst ebenfalls das LO-Prinzip beim Revidieren räumlicher Überzeugungen. Zu diesen Hypothesen habe ich jeweils ein Experiment durchgeführt. In *Experiment 6* habe ich ein künstliches Gruppensetting erzeugt, indem ich die Probanden habe glauben lassen, dass sie über einen Computer die Antworten von drei anderen Gruppenmitgliedern erhalten. Die Probanden dachten, dass diese anderen Personen real existieren und ebenfalls an der Aufgabe arbeiteten. So konnte gezielt manipuliert werden, wie viele Probanden sich entweder für die LO- oder RO-Revision entschieden haben und es konnte hierdurch ein direkter Konformitätsdruck für die Probanden erzeugt werden. Die Methodik von *Experiment 7* ist sehr stark an die Experimente von Wolf et al. (2012) angelehnt. Ich habe den Probanden die Prämissen in Form einer verbal-schriftlichen Unterhaltung von zwei (unterschiedlich vertrauenswürdigen) Berufsgruppen präsentiert. Das heißt, dass sie nicht direkt in einer Gruppe interagierten, sondern vielmehr die Unterhaltung einer anderen Dyade gewissermaßen „beobachteten“ und aufgrund dieser Informationen eine Revisionsentscheidung fällten. Somit konnte ich die Konformität in Experiment 6 und das Quellenprinzip in Experiment 7 manipulieren und prüfen, ob diese das in der Forschung – ohne sozialen Kontext – sehr robuste LO-Prinzip beeinflussen.

2.4 Hypothesen

Ziel der vorliegenden Arbeit ist es also, die unterschiedlichen Phasen beim kollaborativen räumlichen Schlussfolgern und der Überzeugungsrevision mit Gruppen zu testen. Dabei habe ich in drei experimentellen Teilen insgesamt sieben Experimente durchgeführt. Die Befunde dieser Experimente werden in den drei nachfolgenden experimentellen Teilen (Kapitel 3: Kollaborative Konstruktion, Kapitel 4: Kollaborative Variation und Kapitel 5: Soziale Revision) berichtet und anschließend in Kapitel 6 einer allgemeinen Diskussion in eine Theorie zum *kollaborativen räumlichen Schlussfolgern und Überzeugungsrevision* integriert.

Zusammenfassend die Hypothesen der vorliegenden Arbeit:

Experimenteller Teil 1 – Kollaborative Konstruktion mentaler Modelle

Dyaden sind in der Lage, gemeinsam geteilte mentale Modelle eines Problems zu konstruieren. Dabei unterliegen sie dem Kontinuitätseffekt, machen aber weniger Fehler als Individuen. Außerdem wird in der Dyade beim Teilen der individuellen mentalen Modelle häufiger die Annotation über eine Diskontinuität verbalisiert als bei Individuen.

Experimenteller Teil 2 – Kollaborative Variation mentaler Modelle

Dyaden ziehen mehr valide Konklusionen als Individuen. Dabei berücksichtigen sie häufiger mehrere Modelle und profitieren somit am meisten bei indeterminierten Problemen. Die allgemeine Leistung der Dyaden kann am besten durch das *truth wins* Schema vorhergesagt werden.

Experimenteller Teil 3 – Soziale Revision mentaler Modelle

Die räumliche Überzeugungsrevision wird durch soziale Faktoren wie die Konformität und dem Quelleneffekt beeinflusst. Konkret zeigt sich dies darin, dass das LO-Prinzip seltener gezeigt wird, wenn die Majorität gegen das LO-Prinzip ist oder die Aussage (Prämisse) einer wenig vertrauenswürdigen Quelle dem LO-Prinzip entspricht.

Kapitel 3: Experimenteller Teil 1 - Kollaborative Konstruktion mentaler Modelle

In diesem Kapitel habe ich in zwei Experimenten getestet, ob Dyaden in der Lage sind, gemeinsam eine Repräsentation der Prämissen zu konstruieren. Im ersten Experiment habe ich getestet, ob Dyaden dem gleichen Kontinuitätseffekt beim Konstruieren geteilter mentaler Modelle wie Individuen unterliegen. Dazu habe ich den Probanden (Individuen oder Dyaden) räumliche Prämissen in unterschiedlicher Reihenfolge (kontinuierlich, semi-kontinuierlich und diskontinuierlich) präsentiert. Ich habe angenommen, dass auch Dyaden den Kontinuitätseffekt zeigen. Weiterhin ging ich davon aus, dass Dyaden mehr korrekte Modelle konstruieren als Individuen. Im zweiten Experiment habe ich getestet, ob Personen in einer Dyade die Annotation über die Diskontinuität mit ihrem Partner teilen. Dazu habe ich das Experiment so konstruiert, dass eine der beiden Personen immer zwei Prämissen präsentiert bekam und allein ein Modell konstruieren musste. Ich ging davon aus, dass Personen, die gemeinsam in der Dyade arbeiteten, häufiger die Annotation bei diskontinuierlichen Problemen verbalisieren als Individuen. Außerdem habe ich auch in dem zweiten Experiment erwartet, dass Dyaden mehr korrekte Modelle konstruieren als Individuen.

Zusammenfassend die drei Hypothesen dieses Kapitels:

Hypothese 1: Dyaden zeigen den Kontinuitätseffekt beim Konstruieren geteilter mentaler Modelle.

Hypothese 2: Dyaden konstruieren mehr korrekte Modelle als Individuen.

Hypothese 3: Dyaden verbalisieren bei diskontinuierlichen Problemen häufiger die Annotation über die Diskontinuität als Individuen.

Die erste und zweite Hypothese habe ich mit dem ersten Experiment getestet. Mit dem zweiten Experiment habe ich die zweite und die dritte Hypothese geprüft. Dabei habe ich alle Experimente mit dem Programm OpenSesame (Version 3.1, Mathôt, Schreij, & Theeuwes, 2012) programmiert und durchgeführt. Die zufällige Zuordnung der Probanden wurde mit der Randomisierungssoftware von randomizer.org (Version 4, Urbaniak & Plous, 2013) durchgeführt. Die Aufbereitung der Daten und die Erstellung der Diagramme für alle Experimente erfolgten mit Microsoft Excel (Version 2016). Die inferenzstatistische Auswertung der Daten erfolgte mit der Statistiksoftware SPSS (Version 24, IBM Corp.,

2016). Ein wichtiges Maß für die Leistung sind die relativen Häufigkeiten valider Konklusionen oder korrekt konstruierter Modelle (z. B. $.60 = 60\%$). Da die relativen Häufigkeiten zwischen 0 und 1 eingeschlossen sind, können sie keiner Normalverteilung folgen. Da die Normalverteilung eine wichtige Annahme für die varianzanalytische Auswertung ist (Bortz & Schuster, 2010, S. 129), habe ich mich für eine Anpassung entschieden. Um die Daten approximativ einer Normalverteilung anzupassen, habe ich für alle relativen Häufigkeiten (p) eine Angular-Transformation $[(\sin^{-1}(\sqrt{p}))]$ durchgeführt. Für Antwort- und Lesezeiten ist dies nicht notwendig, da sich intervallskalierte Daten nach dem zentralen Grenzwerttheorem bei einer ausreichend großen Stichprobe einer Normalverteilung annähern (Bortz & Schuster, 2010, S. 86). Somit konnten für alle abhängigen Maße parametrische Verfahren durchgeführt werden.

Die vorliegenden Experimente sind im Einklang mit den Standards der aktuellsten Version der Deklaration von Helsinki (World Medical Association, 2013) und wurden von einer lokalen Ethikkommission geprüft. Eine Ausnahme ist dabei, dass für die vorliegenden Experimente keine Pre-Registrierung nach §35 der Deklaration durchgeführt wurde.

3.1 Experiment 1 – Kollaborative Konstruktion mentaler Modelle

Im ersten Experiment dieses Kapitels habe ich den Kontinuitätseffekt bei Dyaden untersucht. Die Methodik dieses Experimentes ist angelehnt an das Experiment 1 von Nejasmic et al. (2015). Diese haben ebenfalls den Kontinuitätseffekt, allerdings bei Individuen, getestet. Deshalb habe ich ihre Aufgaben und den Ablauf übernommen, um es beim Denken von Dyaden zu testen.

3.1.1 Methoden

3.1.1.1 Stichprobe

Die Stichprobe wurde mit Hilfe des Rundmail-Systems der Justus-Liebig-Universität akquiriert. Dabei wurden keine expliziten Ausschlusskriterien vordefiniert. An der Untersuchung nahmen $N = 40$ Versuchspersonen (Vpn) (29 weiblich) teil. Das mittlere Alter lag bei $M = 24.1$ ($SD = 7.7$) Jahren und reichte von $MIN = 18$ bis $MAX = 65$ Jahre. Die Hälfte der Stichprobe arbeitete dabei zu zweit (Dyade Bedingung), die andere Hälfte alleine (Individuum Bedingung). Die Zuordnung zu der Bedingung war dabei randomisiert. Für den Arbeitsaufwand erhielten die Vpn sogenannte Versuchspersonenstunden, die sie im Rahmen ihres Studiums benötigen. Als zusätzlicher Anreiz wurden unter allen Teilnehmern zwei Amazon-Gutscheine im Wert von je 25€ verlost.

3.1.1.2 Material

Das Material enthielt 30 relational räumliche Probleme. Jedes Problem bestand aus drei Prämissen, welche die Relation von insgesamt vier Termen beschrieb (vier-Term Problem). Den Termen wurde zufällig jeweils ein Objekt aus der Kategorie „Obst“ (Apfel, Birne, Mango, Orange, Kiwi) oder „Werkzeug“ (Hammer, Meißel, Säge, Feile, Bohrer) zugeordnet. Für diese Objektkategorien bestehen keine natürlichen Relationen, wie es bei anderen Objekten der Fall sein würde (z. B. beim Gesicht, bei welchem man erwartet, dass die Augen über dem Mund sind). Deshalb kann die Verwendung von Obst und Gemüse gewährleisten, dass kein Hintergrundwissen in die Konstruktion des mentalen Modells einfließt. Die Objekte für die beiden Kategorien wurden so gewählt, dass kein Term den gleichen Anfangsbuchstaben besaß. Ein Problem enthielt nur Objekte aus jeweils einer Kategorie, sodass sie entweder nur die Relationen von Obst oder von Werkzeugen beschreiben konnten. Von den 30 Problemen bestanden je zehn aus einer kontinuierlichen, semi-kontinuierlichen sowie diskontinuierlichen Reihenfolge der Prämissen. Die eine Hälfte dieser zehn Aufgaben enthielt Obst, die andere Hälfte Werkzeuge als Terme. Da die Arbeitsrichtung (links vs. rechts) einen Einfluss auf die Konstruktion und Schwierigkeit der Probleme hat (Nejasmic et al., 2015), habe ich mich entschieden, ausschließlich das Prädikat „links von“ für sämtliche Prämissen zu verwenden. Tabelle 3 stellt die Aufgaben mit ihren verschiedenen Reihenfolgen dar.

Tabelle 3

Verschiedene Prämissenreihenfolgen, die für das Experiment verwendet wurden

	Reihenfolge		
	Kontinuierlich	Semi-Kontinuierlich	Diskontinuierlich
1. Prämisse	A r ₁ B	B r ₂ C	C r ₃ D
2. Prämisse	B r ₂ C	C r ₃ D	A r ₁ B
3. Prämisse	C r ₃ D	A r ₁ B	B r ₂ C
Modell	ABCD		

Anmerkung: Den Termen A bis D wurden die Objekte aus der Kategorie „Obst“ oder „Werkzeug“ zugeordnet. Das Prädikat (r) war immer „links von“.

3.1.1.3 Ablauf

Die Vpn haben entweder alleine oder zu zweit teilgenommen und das gesamte Experiment fand an einem handelsüblichen Computer statt. Wenn die Vpn in der Dyade teilnahmen, saßen die beiden Vpn gemeinsam vor einem Monitor. Am Anfang des Experiments erhielten die

Vpn Informationen zum Datenschutz und der Dauer des Experimentes. Danach unterschrieben die Vpn eine Einverständniserklärung. Anschließend wurde die Instruktion auf dem Monitor präsentiert. Die Vpn wurden darüber informiert, dass sie drei Prämissen in sequentieller (nacheinander) Reihenfolge präsentiert bekommen. Dabei konnten sie mit Druck auf die „Leertaste“ selbstständig entscheiden, wann sie zu der nächsten Prämisse fortfahren wollten. Wenn die Vpn die letzte Prämisse gesehen und die Leertaste gedrückt hatten, wurde ihnen ein Antwortfeld präsentiert. Ihre Aufgabe war es, die Anfangsbuchstaben der Objekte in der Reihenfolge in das Antwortfeld zu schreiben, wie sie es sich zuvor konstruiert hatten (Beispiel: ABMK für Apfel Birne Mango Kiwi). Gemessen wurden jeweils die Zeit, welche die Vpn mit dem Lesen oder Besprechen jeder einzelnen Prämisse verbracht hatten sowie die korrekt eingetragenen Modelle. Wenn die Vpn in einer Dyade arbeiteten, wurde ihnen auf einer zusätzlichen Seite gesagt, dass sie unbedingt gemeinsam zu einer Lösung des Problems kommen sollten und sich so viel Zeit zum Diskutieren lassen konnten, wie sie wollten. Außerdem wurde ihnen gesagt, dass sie jederzeit – also auch während der Präsentation der Prämissen – miteinander reden durften. Damit sollte sichergestellt werden, dass die beiden Personen gemeinsam auf eine Lösung kamen. Nach jedem Durchgang erschien ein Pausenfenster und die Vpn konnten selbstständig entscheiden, wann sie fortfahren. Bevor das eigentliche Experiment startete, bearbeiteten alle Vpn drei Übungsaufgaben. Dabei entsprachen die Aufgaben jeweils einem kontinuierlichen, einem semi-kontinuierlichen und einem diskontinuierlichen Problem. Den Vpn wurde gesagt, dass sie sich so viel Zeit für die Aufgaben nehmen konnten, wie sie benötigten, aber dabei trotzdem versuchen sollten, schnellstmöglich zu antworten. Bei den Übungsaufgaben konnten sie dem Experimentalleiter noch Fragen stellen. Während des Hauptexperimentes war dies jedoch nicht mehr möglich. Alle 30 Probleme des Hauptexperimentes wurden in randomisierter Reihenfolge präsentiert. Nachdem die Vpn alle Probleme bearbeitet hatten, wurden sie über das Ziel des Experimentes aufgeklärt und für ihren Aufwand mit Versuchspersonenstunden entschädigt. Das Experiment dauerte insgesamt eine Stunde.

3.1.1.4 Design

Das Experiment folgt einem 2×3 gemischten Design mit dem Zwischensubjektfaktor „Gruppe“ (Individuum \times Dyade) und dem Innersubjektfaktor „Kontinuität“ (kontinuierlich \times semi-kontinuierlich \times diskontinuierlich). Die abhängigen Variablen sind die Lesezeit der Prämissen und korrekt konstruierte Modelle. Das experimentelle Design ist als Überblick mit der Anzahl der Probleme in Tabelle 4 dargestellt.

Tabelle 4

Experimentelles Design mit $N = 40$ Versuchspersonen. Die abhängigen Variablen sind die Lesezeit der Prämissen und korrekt konstruierte Modelle

Gruppe (Zwischensubjektfaktor)	Kontinuität (Innersubjektfaktor)		
	Kontinuierlich	Semi- Kontinuierlich	Diskontinuierlich
	Individuen	10 Probleme	10 Probleme
Dyaden	10 Probleme	10 Probleme	10 Probleme

3.1.2 Ergebnisse

Für die Auswertung der Daten wurden aus den $n = 20$ individuellen Datensätzen zufällig jeweils zwei Datensätze ausgewählt und zusammengefasst. Somit wurde aus zwei Individuen eine statistische Dyade geformt. Somit konnte ich die Anzahl an Beobachtungen für beide Bedingungen gleich halten. Das bedeutet, dass für die Auswertung $n = 10$ Datensätze der Individuen mit $n = 10$ Datensätzen der Dyaden verglichen wurden. Es wurden keine Datensätze aus der Auswertung ausgeschlossen. Da ich keinen Vergleich zwischen Individuen und Dyaden bei den Lesezeiten geplant hatte, habe ich für diese Lesezeiten ausschließlich die Datensätze der Dyaden verwendet. Bei den korrekt konstruierten Modellen habe ich jedoch die Gruppen miteinander verglichen.

3.1.2.1 Lesezeit der Prämissen

Deskriptiv zeigt sich bei den Dyaden für die erste Prämisse kaum ein Unterschied in den Lesezeiten (kontinuierlich: $M = 4.59s$, $SD = 1.37s$; semi-kontinuierlich: $M = 4.37s$, $SD = 1.88s$; diskontinuierlich: $M = 4.36s$, $SD = 1.47s$). Bei der zweiten Prämisse zeigen sich ähnliche Lesezeiten, wobei diskontinuierliche bereits deskriptiv etwas länger verarbeitet wurden (kontinuierlich: $M = 5.42s$, $SD = 1.93s$; semi-kontinuierlich: $M = 5.06s$, $SD = 1.98s$; diskontinuierlich: $M = 6.91s$, $SD = 2.30s$). Die wichtige Prämisse für den Kontinuitätseffekt ist jedoch die dritte. Bei der dritten Prämisse zeigt sich deskriptiv, dass Dyaden für kontinuierliche Probleme ($M = 5.90s$, $SD = 1.91s$) nicht mehr Zeit benötigten als für semi-kontinuierliche ($M = 6.22s$, $SD = 2.50s$). Bei diskontinuierlichen Problemen ($M = 11.29s$, $SD = 4.45s$) benötigten Dyaden für die Verarbeitung deskriptiv deutlich länger als bei

kontinuierlichen und semi-kontinuierlichen Problemen. Die Lesezeiten der Prämissen für Dyaden sind in Abbildung 1 zur Veranschaulichung dargestellt.

Für die Auswertung der Lesezeiten habe ich die Prämissen als Faktor mit drei Stufen definiert und eine 2×3 Varianzanalyse mit Messwiederholung mit den Innersubjektfaktoren „Prämisse“ (erste \times zweite \times dritte Prämisse) und „Kontinuität“ (kontinuierlich \times semi-kontinuierlich \times diskontinuierlich) durchgeführt. Die Analyse ergab einen signifikanten Haupteffekt für den Faktor „Kontinuität“ ($F(2, 18) = 9.54, p = .001, \eta_p^2 = .515$) und ebenfalls einen Haupteffekt für den Faktor „Prämisse“ ($F(2, 18) = 50.29, p < .001, \eta_p^2 = .848$). Für meine Auswertung war die signifikante Interaktion „Kontinuität \times Prämisse“ ($F(4, 36) = 8.76, p = .002, \eta_p^2 = .494$, korrigiert nach Greenhouse-Geisser) am wichtigsten. Um die Interaktion genauer zu untersuchen, habe ich post hoc t -Tests für verbundene Stichproben durchgeführt. Um einer Alpha-Fehler Kumulation durch multiples Testen entgegenzuwirken, habe ich das Alpha-Niveau nach Bonferroni korrigiert ($\alpha = .0167$). Für die erste und zweite Prämisse zeigten sich keine Unterschiede in Abhängigkeit der Kontinuität ($p > .0167$). Bei der dritten Prämisse war die Lesezeit bei kontinuierlichen Problemen signifikant kürzer als bei diskontinuierlichen Problemen ($t(9) = 3.65, p = .005, d = 1.57$). Semi-kontinuierliche Probleme wurden ebenfalls signifikant schneller von den Dyaden verarbeitet als diskontinuierliche Probleme ($t(9) = 4.04, p = .003, d = 1.41$). Der Vergleich zwischen kontinuierlichen und semi-kontinuierlichen Problemen ergab bei der dritten Prämisse keinen signifikanten Unterschied ($t(9) = 0.36, p = .73, d = 0.14$).

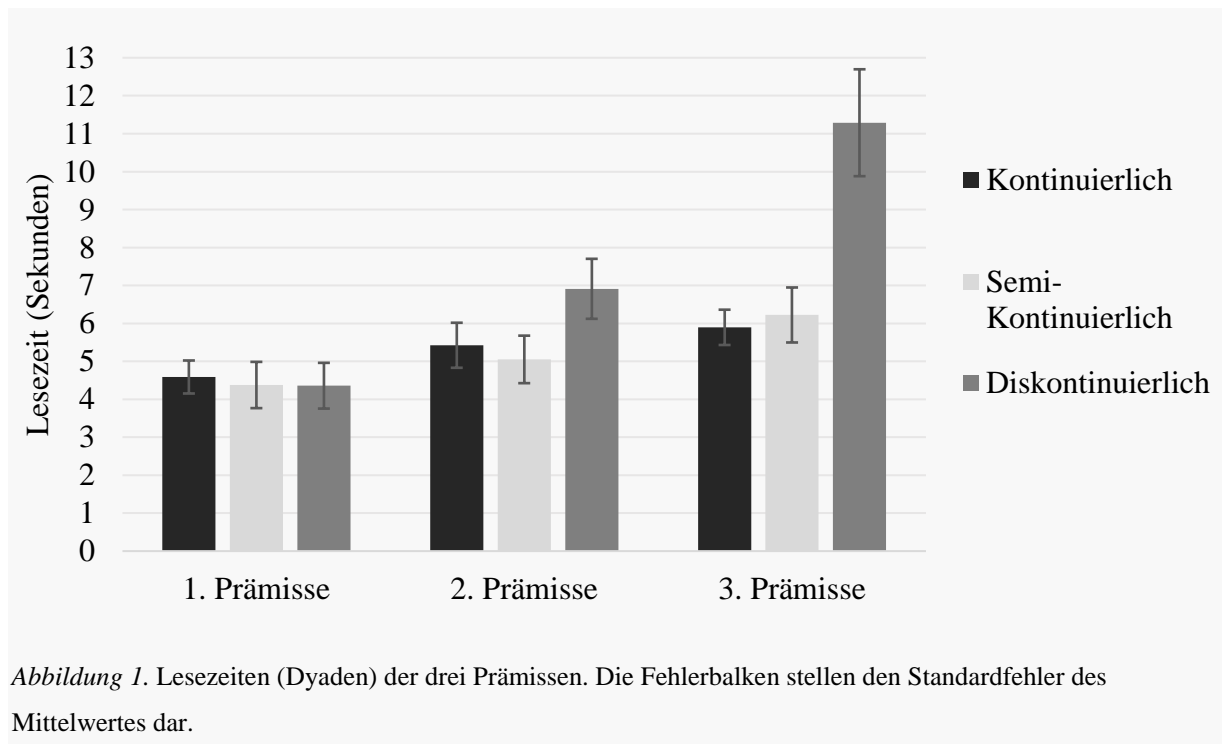


Abbildung 1. Lesezeiten (Dyaden) der drei Prämissen. Die Fehlerbalken stellen den Standardfehler des Mittelwertes dar.

3.1.2.2 Korrekte Modelle

Deskriptiv zeigt sich, dass bei kontinuierlichen Problemen ($M = 93.25\%$, $SD = 10.04\%$) ähnlich viele korrekte Modelle konstruiert wurden wie bei semi-kontinuierlichen Problemen ($M = 94.25\%$, $SD = 09.63\%$). Diskontinuierliche Probleme waren mit 88.25% ($SD = 14.26\%$) deskriptiv am schwierigsten. Außerdem zeigen die deskriptiven Daten, dass Dyaden bei allen Problemen mehr korrekte Modelle konstruierten als Individuen. Bei kontinuierlichen Problemen erreichten Dyaden 96.00% ($SD = 5.16\%$) korrekte Modelle (Individuen: $M = 90.50\%$, $SD = 13.01\%$). Bei semi-kontinuierlichen Problemen konnten Dyaden 99.00% ($SD = 3.16\%$) korrekte Modelle konstruieren (Individuen: $M = 89.50\%$, $SD = 11.66\%$). Bei diskontinuierlichen Problemen erreichten Dyaden 93.00% ($SD = 6.75\%$) und Individuen 83.50% ($SD = 18.27\%$) korrekte Modelle. Zur besseren Veranschaulichung sind die deskriptiven Daten in Abbildung 2 dargestellt.

Zur inferenzstatistischen Überprüfung dieser Unterschiede habe ich eine 2×3 Varianzanalyse mit Messwiederholung mit dem Zwischensubjektfaktor „Gruppe“ (Individuum \times Dyade) und dem Innersubjektfaktor „Kontinuität“ (kontinuierlich \times semi-kontinuierlich \times diskontinuierlich) durchgeführt. Es ergab sich ein signifikanter Haupteffekt für den Faktor „Kontinuität“ ($F(2, 36) = 3.49$, $p = .041$, $\eta_p^2 = .162$). Keine der post-Hoc t -Tests für verbundene Stichproben konnten jedoch das nach Bonferroni korrigierte Niveau ($\alpha = .017$) unterschreiten (kontinuierlich vs. semi-kontinuierlich, $t(19) = -.90$, $p = .379$, $d = 0.10$; kontinuierlich vs. diskontinuierlich, $t(19) = 1.71$, $p = .104$, $d = 0.40$; semi-kontinuierlich vs.

diskontinuierlich, $t(19) = 2.49$, $p = .022$, $d = 0.50$). Der für meine Fragestellung interessanteste Haupteffekt für den Faktor „Gruppe“ erreichte Signifikanz ($F(1, 18) = 5.57$, $p = .030$, $\eta_p^2 = .236$). Dyaden konnten insgesamt mehr korrekte Modelle konstruieren als Individuen. Die Interaktion „Gruppe×Kontinuität“ konnte nicht nachgewiesen werden ($F(2, 36) = 1.11$, $p = .341$, $\eta_p^2 = .058$).

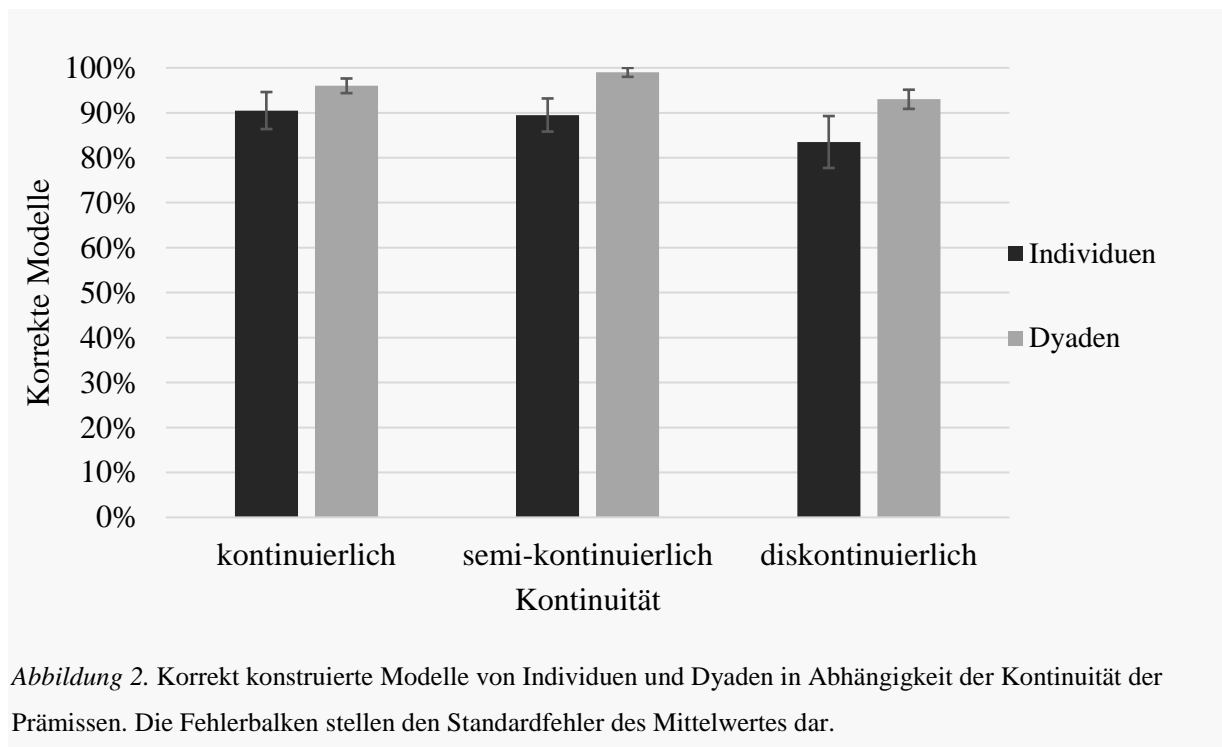


Abbildung 2. Korrekt konstruierte Modelle von Individuen und Dyaden in Abhängigkeit der Kontinuität der Prämissen. Die Fehlerbalken stellen den Standardfehler des Mittelwertes dar.

3.1.3 Diskussion

Dyaden benötigten bei kontinuierlichen Problemen im Vergleich zu diskontinuierlichen Problemen fünf Sekunden weniger Zeit für die Verarbeitung der dritten Prämisse. Auch bei semi-kontinuierlichen Problemen verarbeiteten die Dyaden die dritte Prämisse um fünf Sekunden schneller als bei diskontinuierlichen Problemen. Diese Befunde stützen meine erste Hypothese, dass *Dyaden den Kontinuitätseffekt beim Konstruieren geteilter mentaler Modelle zeigen*. Wenn die Prämissenreihenfolge diskontinuierlich ist, benötigen Dyaden mehr Zeit, um die dritte Prämisse zu verarbeiten. Die Befunde zu den korrekt konstruierten Modellen zeigen, dass Dyaden über alle Probleme hinweg 8 % mehr korrekte Modelle konstruieren konnten als Individuen. Diese Daten stützen meine zweite Hypothese, dass *Dyaden mehr korrekte mentale Modelle konstruieren als Individuen*. Dyaden zeigen also – analog zu Individuen – den Kontinuitätseffekt, konstruieren aber mehr korrekte Modelle als Individuen. Das bedeutet, dass Dyaden (1) in der Lage sind, gemeinsam ein Modell eines räumlich-relationalen Problems zu konstruieren und (2) die Leistung von Individuen übertreffen können. Diese

Befunde sind eine sehr wichtige Grundlage für die Erforschung des kollaborativen räumlichen Schlussfolgerns. Wären Gruppen nicht in der Lage, gemeinsam eine Repräsentation des Problems zu konstruieren, könnten sie keine validen Konklusionen ziehen.

Durch dieses Experiment können wir sehen, dass Dyaden ihr geteiltes mentales Modell sehr ähnlich zum individuellen Denken konstruieren. Dies entspricht der Theorie der Informationsverarbeitung von Gruppen nach Hinsz et al. (1997). Diese besagt, dass das geteilte mentale Modell das Resultat aus den individuellen mentalen Modellen ist. Würde der Konstruktionsprozess bei Gruppen unabhängig vom individuellen Denken sein, hätte sich der Kontinuitätseffekt nicht zeigen müssen. So hätte die Dyade den Effekt zum Beispiel umgehen können, indem eine Person eine Prämisse und die andere Person die anderen Prämissen im Arbeitsgedächtnis hält. Die Integration der Prämissen in ein Modell hätten sie dann erst bei Erscheinen des Antwortfensters durchführen können. Dies hätte zur Folge, dass die Verarbeitungszeit der einzelnen Prämissen extrem gering wäre und der Kontinuitätseffekt bei den Dyaden sich nicht in den Daten zeigen ließe. Somit sprechen meine Daten dafür, dass Dyaden – analog zum individuellen Denken – die Informationen aus den Prämissen schrittweise in ein geteiltes mentales Modell integrieren.

Die Theorie nach Hinsz et al. (1997) nimmt außerdem an, dass Gruppen Informationen besser verarbeiten als Individuen. Obwohl die Dyaden in meinem Experiment im Vergleich zu Individuen bei allen Problemen mehr korrekte Modelle konstruieren, hätte man erwarten können, dass sie durch die Konstruktion eines geteilten mentalen Modells besonders bei den schwierigen (diskontinuierlichen) Problemen profitieren. Dies ist aber nicht der Fall. Dyaden haben, genau wie bei semi-kontinuierlichen Problemen, 10 % mehr korrekte Modelle konstruiert als Individuen. Ich nehme an, dass dieser Befund auf einen allgemeinen Deckeneffekt zurückzuführen ist. Alle Vpn – Individuen und Dyaden – haben sehr viele korrekte Modelle konstruiert. Somit scheinen auch Individuen die Probleme sehr leicht lösen zu können und ein Unterschied in der Informationsverarbeitung kann sich nur geringfügig zeigen. Um zu prüfen, ob Dyaden mehr Informationen verarbeiten als Individuen sind Probleme nötig, die eine höhere relationale Komplexität (Halford et al., 1989) aufweisen.

Insgesamt stützen meine Befunde die Annahme, dass das geteilte mentale Modell auf der Kombination der individuellen Modelle basiert. Dabei kann dieses Experiment aber nichts darüber aussagen, wie diese individuellen Modelle kombiniert werden. Deshalb soll im nächsten Experiment getestet werden, wie Personen beim Kollaborieren in einer Dyade ihre (vorläufig) konstruierten mentalen Modelle miteinander austauschen.

3.2 Experiment 2 – Kollaborative Konstruktion und Teilen mentaler Modelle

Im zweiten Experiment dieses Kapitels habe ich untersucht, ob Personen, die zunächst individuell ein mentales Modell konstruiert haben, die Annotation bei diskontinuierlichen Problemen ihrem Partner mitteilen. Die Methodik dieses Experimentes ist sehr ähnlich zu dem ersten Experiment. Ich habe das Experiment jedoch so angepasst, dass eine Person in der Dyade immer zwei Prämissen präsentiert bekam. So konnte ich sicherstellen, dass immer nur eine Person ein vorläufiges Modell konstruiert und dieses mit dem Partner teilt.

3.2.1 Methoden

3.2.1.1 Stichprobe

Die Stichprobe wurde mit Hilfe des Rundmail-Systems der Justus-Liebig-Universität akquiriert. Dabei wurden keine expliziten Ausschlusskriterien vordefiniert. An der Untersuchung nahmen $N = 51$ Vpn (40 weiblich) teil. Das mittlere Alter lag bei $M = 22.6$ ($SD = 3.7$) Jahren und reichte von $MIN = 18$ bis $MAX = 37$ Jahre. Von der Stichprobe arbeiteten $n = 36$ zu zweit (18 Dyaden), die anderen $n = 15$ Personen alleine (Individuum). Die Zuordnung zu der Bedingung war dabei randomisiert. Für den Arbeitsaufwand erhielten die Vpn Versuchspersonenstunden. Als zusätzlicher Anreiz wurden unter allen Teilnehmern fünf Amazon-Gutscheine im Wert von je 20 € verlost.

3.2.1.2 Material

Das Material war sehr ähnlich zu dem aus Experiment 1. Es wurden insgesamt 16 vier-Term Probleme mit Objekten entweder aus der Kategorie „Obst“ (Apfel, Birne, Mango, Kiwi, Orange) oder „Werkzeug“ (Hammer, Meißel, Säge, Bohrer, Feile) verwendet. Für das Prädikat wurde in diesem Experiment „links von“ und „rechts von“ verwendet. Dies sollte sicherstellen, dass die Probanden nicht schon mit Präsentation der zweiten Prämisse das zu erwartende gesamte Modell antizipieren konnten. In diesem Experiment war nur der Vergleich zwischen kontinuierlich und diskontinuierlich präsentierten Prämissen von Interesse. Deshalb habe ich auf die semi-kontinuierliche Prämissenreihenfolge verzichtet. Das bedeutet, dass jeweils acht kontinuierliche und acht diskontinuierliche Probleme präsentiert wurden. Von diesen waren wieder die Hälfte der Probleme mit den Objekten der Kategorie „Obst“ und die andere Hälfte mit Objekten der Kategorie „Werkzeug“. Tabelle 5 stellt das Schema für das Aufgabenmaterial dar.

Tabelle 5

Aufgabenmaterial mit den verschiedenen Prämissenreihenfolgen

	Reihenfolge	
	Kontinuierlich	Diskontinuierlich
1. Prämisse	A r ₁ B	C r ₃ D
2. Prämisse	B r ₂ C	A r ₁ B
3. Prämisse	C r ₃ D	B r ₂ C
Resultierendes Modell	ABCD	

Anmerkung: Den Termen A bis D wurden die Objekte aus der Kategorie „Obst“ oder „Werkzeug“ zugeordnet. Das Prädikat (r) war „links von“ oder „rechts von“.

3.2.1.3 Ablauf

Das gesamte Experiment fand an einem handelsüblichen Computer statt. Am Anfang des Experiments erhielten die Vpn Informationen zum Datenschutz und der Dauer der Untersuchung. Danach unterschrieben die Vpn eine Einverständniserklärung. Anschließend wurde die Instruktion präsentiert. Der Ablauf für die Bedingung „Individuen“ war sehr ähnlich zu dem Ablauf aus Experiment 1. Die Probanden bekamen auf einem Computerbildschirm sequentiell jeweils eine Prämisse präsentiert. Sie konnten in diesem Experiment aber nicht selbstständig entscheiden, wann sie zu der nächsten Prämisse fortfahren würden. Die Präsentationszeit für jede Prämisse wurde auf 20 Sekunden festgesetzt, um die Ergebnisse besser mit denen der Dyade vergleichbar zu machen. Das heißt, dass das Experiment automatisch nach 20 Sekunden mit der nächsten Prämisse fortfuhr. Nach der zweiten Prämisse wurden die Vpn aufgefordert, ihr vorläufig konstruiertes Modell laut aufzusagen. Die Verbalisierung des Modells wurde von dem Experimentalleiter notiert und in zwei Kategorien eingeordnet. Alle Fälle, bei denen sie explizit äußerten, dass die Relation der vier Objekte nicht eindeutig war (z. B. „Apfel Birne und Mango Kiwi“ oder „Apfel Birne und die anderen Objekte sind Mango Kiwi“) wurden in die Kategorie „Annotation berücksichtigt“ eingeordnet. Zu „Annotation nicht berücksichtigt“ zählten alle Durchgänge, bei denen die Vpn ein zusammenhängendes Modell (z. B. „Apfel Birne Mango Kiwi“) oder den exakten Wortlaut der Prämissen verbalisierten. Nachdem die Vpn das Modell aufgesagt hatte, konnte sie mit der Leertaste zur dritten Prämisse fortfahren. Wieder nach 20 Sekunden erschien ein Antwortfenster, in dem die Vpn das Modell in Form der Anfangsbuchstaben der Objekte (z. B. ABMK) eintragen musste. Nach dem Antwortfenster konnten die Vpn pausieren und selbstständig entscheiden, wann sie fortfahren.

Der Ablauf für die Bedingung „Dyade“ war etwas anders konstruiert. In diesem Fall saßen die Probanden vor zwei Monitoren, die von einem Sichtschutz getrennt waren. So konnte jede Vpn nur die Informationen sehen, die auf ihrem Monitor präsentiert wurde. Einer Person wurden die ersten beiden Prämissen gleichzeitig für 40 Sekunden präsentiert. Während dieser Zeit erhielt die andere Person die dritte Prämisse (ebenfalls 40 Sekunden) und die Vpn durften nicht miteinander kommunizieren. Nachdem die 40 Sekunden abgelaufen waren, wurden die Prämissen ausgeblendet und der Experimentalleiter entfernte den Sichtschutz. Die Vpn waren instruiert, die Informationen miteinander auszutauschen und gemeinsam ein zusammenhängendes Modell zu konstruieren. Dabei notierte der Experimentalleiter die Verbalisierung des vorläufigen Modells von der Person, welche die ersten zwei Prämissen präsentiert bekam. Diese wurden der gleichen Kategorie (mit Annotation vs. ohne Annotation) wie schon in der Bedingung „Individuum“ zugeordnet. Während des Experimentes wurde pseudo-randomisiert entschieden, welcher der beiden Personen zwei und welche eine Prämisse präsentiert bekam. Am Ende des Experimentes hat jede Vpn jeweils in 50 % der Fälle die ersten beiden oder die dritte Prämisse präsentiert bekommen. Nachdem die Dyade gemeinsam das konstruierte Modell in das Antwortfenster notierte (z. B. ABMK), konnte sie pausieren. Der Experimentalleiter stellte den Sichtschutz zurück zwischen die Monitore und die Vpn konnten gemeinsam entscheiden, wann sie fortfuhren. Vor dem Hauptexperiment bearbeiteten alle Vpn (Individuen und Dyaden) zwei Übungsaufgaben. Davon folgte eine der kontinuierlichen und eine der diskontinuierlichen Prämissenreihenfolge. Nachdem die Vpn alle 16 Probleme bearbeitet hatten, wurden sie über Ziel des Experimentes aufgeklärt und für ihren Aufwand mit Versuchspersonenstunden entschädigt. Das Experiment dauerte insgesamt eine Stunde.

3.2.1.4 Design

Das Experiment folgt einem 2×2 gemischten Design mit dem Zwischensubjektfaktor „Gruppe“ (Individuum \times Dyade) und dem Innersubjektfaktor „Kontinuität“ (kontinuierlich \times diskontinuierlich). Abhängige Variablen sind die Verbalisierung der Annotation und die korrekt konstruierten Modelle. Das experimentelle Design ist als Überblick mit der Anzahl der Probleme in Tabelle 6 dargestellt.

Tabelle 6

Experimentelles Design mit $N = 51$ Versuchspersonen. Die abhängigen Variablen sind die Verbalisierung der Annotation und korrekt konstruierte Modelle

Gruppe (Zwischensubjektfaktor)	Kontinuität (Innersubjektfaktor)	
	Kontinuierlich	Diskontinuierlich
Individuen	8 Probleme	8 Probleme
Dyaden	8 Probleme	8 Probleme

3.2.2 Ergebnisse

Vor der Auswertung der Daten habe ich relative Häufigkeiten aus den verbalisierten Annotationen gebildet. Diese geben an, wie häufig eine Person das Modell mit Annotation verbalisiert hat. Von insgesamt $N = 528$ Durchgängen über alle Bedingungen hinweg wurden $n = 21$ (4 %) Verbalisierungen als fehlerhaft (z. B. unverständlich gesprochen) kodiert und aus der Auswertung für die Verbalisierung ausgeschlossen. Für die Analyse der korrekt gebildeten Modelle wurden alle Datensätze ausgewertet. Um die Anzahl an Beobachtungen gleich zu halten, habe ich in diesem Experiment nicht die $n = 15$ Datensätze der Individuen zu einer statistischen Dyade gebildet. Das bedeutet, dass für die Auswertung $n = 15$ Datensätze der Individuen mit $n = 18$ Datensätzen der Dyaden verglichen wurden.

3.2.2.1 Verbalisierung der Annotation

Die kontinuierlichen Probleme erfordern keine Annotation, weshalb diese auch nicht verbalisiert werden konnte. Deshalb reduziert sich die Analyse für die Verbalisierung auf diskontinuierliche Probleme und den Faktor „Gruppe“. Die deskriptiven Daten zeigen, dass Dyaden mit 33.04 % ($SD = 28.79$ %) häufiger die Annotation verbalisiert haben als Individuen ($M = 11.67$ %, $SD = 31.11$ %). Diesen Unterschied konnte ich mit einem t -Test für unabhängige Stichproben untermauern ($t(31) = 2.83$, $p = .025$, $d = .71$).

3.2.2.2 Korrekte Modelle

Die deskriptiven Daten zeigen, dass bei kontinuierlichen Problemen mit 91.67 %, ($SD = 13.13$ %) mehr korrekte Modelle konstruiert wurden als bei diskontinuierlichen Problemen ($M = 73.11$ %, $SD = 27.80$ %). Dyaden konnten bei kontinuierlichen Problemen mehr korrekte

Modelle konstruieren als Individuen (Dyade: $M = 94.44\%$, $SD = 08.81\%$; Individuen: $M = 88.33\%$, $SD = 16.68\%$). Bei diskontinuierlichen Problemen zeigte sich für Dyaden ($M = 72.92\%$, $SD = 25.46\%$) und Individuen ($M = 73.33\%$, $SD = 31.29\%$) eine vergleichbare Leistung. Die Daten sind zur Veranschaulichung in Abbildung 3 dargestellt.

Um die deskriptiven Unterschiede inferenzstatistisch zu prüfen, habe ich eine 2×2 Varianzanalyse mit Messwiederholung mit dem Zwischensubjektfaktor „Gruppe“ (Individuum \times Dyade) und dem Innersubjektfaktor „Kontinuität“ (kontinuierlich \times diskontinuierlich) durchgeführt. Die Analyse ergab keinen signifikanten Haupteffekt für den Faktor „Gruppe“ ($F(1, 31) = 0.26$, $p = .617$, $\eta_p^2 = .008$) und einen signifikanten Haupteffekt für den Faktor „Kontinuität“ ($F(1, 31) = 19.75$, $p < .001$, $\eta_p^2 = .389$). Die Interaktion „Gruppe \times Kontinuität“ ($F(1, 31) = 0.57$, $p = .457$, $\eta_p^2 = .018$) wurde nicht signifikant.

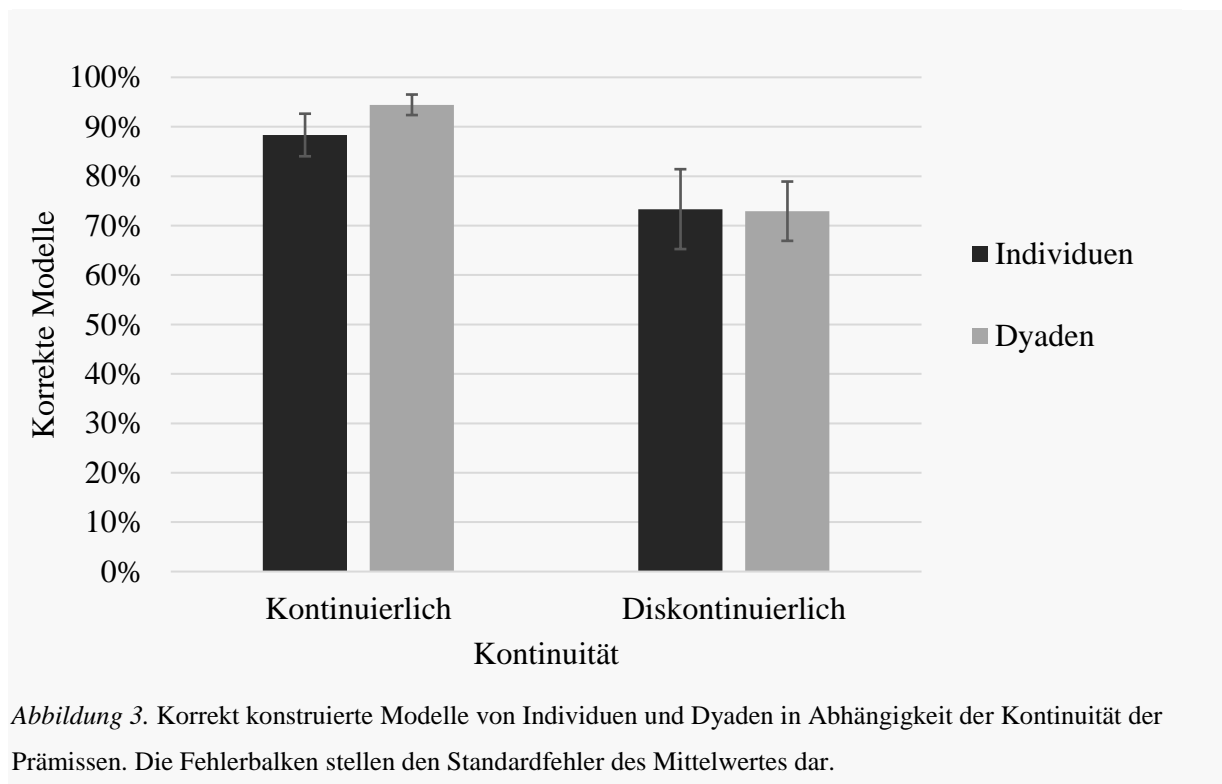


Abbildung 3. Korrekt konstruierte Modelle von Individuen und Dyaden in Abhängigkeit der Kontinuität der Prämissen. Die Fehlerbalken stellen den Standardfehler des Mittelwertes dar.

3.2.3 Diskussion

Ziel dieses Experimentes war zu prüfen, wie Personen ihre mentalen Modelle mit einem Partner in der Dyade teilen. Wenn Personen alleine ein diskontinuierliches Problem lösten und gebeten wurden, ihr vorläufig konstruiertes Modell zu verbalisieren, haben sie in nur 12 % der Fälle die Annotation über diese Diskontinuität berücksichtigt. Wenn die Vpn jedoch in einer Dyade kollaboriert haben, entstand ein anderes Muster. Die Vpn teilten ihrem Partner in 33 % der Fälle mit, dass das Modell vorläufig ist. Diese Daten stützen meine dritte

Hypothese, dass *Dyaden bei diskontinuierlichen Problemen häufiger die Annotation über die Diskontinuität verbalisieren als Individuen*. Die Analyse der korrekt konstruierten Modelle zeigt, dass diskontinuierliche Probleme schwieriger zu lösen sind als kontinuierliche. Dabei machte es keinen Unterschied, ob die Aufgaben von Individuen oder Dyaden bearbeitet wurden. Das bedeutet, dass die vorliegenden Daten gegen meine zweite Hypothese, *Dyaden konstruieren mehr korrekte Modelle als Individuen*, sprechen. Obwohl die Dyaden in einem Drittel der Fälle versucht haben, möglichst viele ihrer individuellen Informationen zu teilen, konnten sie nicht mehr korrekte Modelle konstruieren als Individuen. Dieser Befund stimmt nicht mit den Ergebnissen aus meinem ersten Experiment überein. Es ist davon auszugehen, dass dieser Befund auf die Methodik des vorliegenden Experimentes zurückzuführen ist. So waren die Lesezeiten der Prämissen auf jeweils 20 Sekunden pro Prämisse festgelegt und die Dyaden durften während dieser Zeit nicht kommunizieren. Dies war notwendig, um die Verbalisierung operationalisierbar zu machen. So war sichergestellt, dass nur eine Person ein vorläufiges Modell konstruiert und dieses darauf mit ihrem Partner teilen kann. Dies ist möglicherweise die Ursache dafür, dass die Gruppe keine bessere Leistung erreichen konnte. Wenn beide Personen der Dyade individuell eine Repräsentation des Problems bilden können und dieses in ein geteiltes mentales Modell integrieren, wäre es – analog zu den Befunden aus Experiment 1 – wahrscheinlich, dass die Dyade eine bessere Leistung erreicht.

Personen, die in einer Dyade kollaborativ arbeiteten, haben häufiger die Annotation über eine Diskontinuität verbalisiert als Individuen. Trotzdem haben Dyaden insgesamt bei nur 33 % der diskontinuierlichen Probleme diese Annotation berücksichtigt. Eine Erklärung dafür ist, dass die explizite Nennung der Annotation für die Lösung diskontinuierlicher Probleme zwar hilfreich, aber nicht notwendig war. So ist es möglich, dass eine Person A ein vorläufiges Modell konstruiert und sich die dritte Prämisse von ihrem Partner (Person B) sagen lassen konnte. Daraufhin variierte Person A ihr mentales Modell alleine und kann die Antwort eingeben. Das bedeutet, dass die Leistung bei der Konstruktion nahezu ausschließlich auf den kognitiven Fähigkeiten eines Individuums (Person A) basiert. Diese Interpretation wird durch den Befund, dass Dyaden die gleiche Leistung wie Individuen erzielen, gestützt. Ein weiterer interessanter Punkt ist die Zusammensetzung aller Verbalisierungen der Dyade bei diskontinuierlichen Problemen. In 33 % haben die Personen in der Dyade die Annotation verbalisiert. Wenn man jedoch die Zusammensetzung der Kategorie „Annotation nicht berücksichtigt“ genauer betrachtet zeigt sich, dass in 48 % die Vpn ein zusammenhängendes Modell mit ihrem Partner geteilt haben. Bei 19 % der Durchgänge haben die Vpn sich den genauen Wortlaut der Prämissen gemerkt und ihrem

Partner mitgeteilt. Obwohl es für die Vpn ein leichtes gewesen sein müsste, sich die zwei Prämissen im Wortlaut zu merken und diese dann ihrem Partner mitzuteilen, haben die Vpn die Prämissen in 81 % der Fälle als mentales Modell repräsentiert und dann verbalisiert. Dies ist ein Hinweis darauf, dass das Denken auch im Gruppenkontext am besten durch die Theorie mentaler Modelle beschrieben werden kann.

Zusammenfassend lässt sich sagen, dass uns das Experiment 2 wichtige Aufschlüsse darüber gibt, wie Personen in einer Dyade gemeinsam mit einem Partner denken. Obwohl es für die Lösung nicht relevant war, haben die Personen beim kollaborativen Konstruieren mentaler Modelle versucht, alle Informationen inklusive der Annotation miteinander zu teilen. Im nächsten Abschnitt werde ich die Befunde aus den ersten beiden Experimenten zusammenfassen und zeigen, weshalb diese Experimente einen wichtigen Grundpfeiler für meine weiteren Experimente zum kollaborativen Schlussfolgern bilden.

3.3 Zusammenfassung Kapitel 3

Die Konstruktion mentaler Modelle ist eine wichtige Phase beim Denken. Wenn eine Person oder Gruppe keine korrekte Repräsentation des Problems konstruiert, kann sie auch bei späteren Schlussfolgerungen keine validen Konklusionen ziehen. Deshalb habe ich in diesem Kapitel geprüft, ob Dyaden in der Lage sind, gemeinsam eine Repräsentation aus räumlichen Prämissen zu konstruieren. Ich habe angenommen, dass das geteilte mentale Modell einer Dyade auf der Integration individueller mentaler Modelle basiert. Deshalb habe ich erwartet, dass auch Dyaden dem Kontinuitätseffekt unterliegen und gleichzeitig mehr korrekte Modelle konstruieren. Experiment 1 stützt diese Annahmen. Dyaden brauchten im Vergleich zu kontinuierlichen und semi-kontinuierlichen Problemen deutlich mehr Zeit für die Verarbeitung der dritten Prämisse, wenn die Probleme diskontinuierlich waren.

Experiment 2 konnte uns zeigen, wie in einer Dyade die Informationen untereinander geteilt werden. Wenn Personen zusammen mit einem Partner diskontinuierliche Probleme bearbeiteten, haben sie ihr vorläufig konstruiertes mentales Modell so detailliert wie möglich mit dem Partner geteilt. Das bedeutet, dass sie daran interessiert waren, so viel wie möglich Informationen untereinander auszutauschen. Dieser Befund entspricht der Theorie der Informationsverarbeitung nach Hinsz et al. (1997) und sollte sich deshalb ebenfalls in einer höheren Leistung der Dyaden zeigen. Dies konnte in Experiment 2 jedoch nicht gestützt werden. Die korrekt konstruierten Modelle der Dyaden entsprechen sehr genau der Leistung von Individuen. Eine mögliche Erklärung für diesen Befund ist, dass immer nur eine Person in der Dyade ein vorläufiges Modell konstruieren musste. Somit wurde das geteilte mentale

Modell nicht durch mehrere individuelle Modelle konstruiert, sondern entsprach dem Modell von der Person, welche die ersten beiden Prämissen encodiert hatte. Dies war in Experiment 1 anders. Beide Personen der Dyade konnten alle Prämissen zeitgleich verarbeiten und es zeigte sich, dass die Dyade mehr korrekte Modelle konstruieren konnte als Individuen. Insgesamt waren die Aufgaben zum Kontinuitätseffekt relativ einfach. Die Aufgaben enthielten vier Terme und waren alle auf einer horizontalen (links – rechts) Dimension. In meinem Experiment 1 konnten Individuen 84 % korrekte Modelle konstruieren. Dyaden übertrafen diese Leistung sogar und erreichten 93 %. Auch in Experiment 2 erreichten alle Probanden im Durchschnitt 82 % korrekte Modelle. Dies spricht insgesamt für das Vorhandensein eines Deckeneffektes.

Dieses Kapitel konnte uns also auf der einen Seite zeigen, dass Dyaden sehr gut darin sind, eine Repräsentation der Probleme zu konstruieren. Auf der anderen Seite sind die Probleme zu leicht, um komplexe Denkprozesse beim Schlussfolgern zu untersuchen. So ist zu erwarten, dass Dyaden gerade bei schwierigeren Problemen von der Kollaboration profitieren sollten. Deshalb habe ich im nachfolgenden Kapitel deutlich komplexere Aufgaben verwendet und werde zeigen, wie Dyaden im Vergleich zu Individuen beim Unbestimmtheitseffekt denken und ob sich bei diesen Problemen ein eindeutiger Leistungsunterschied zeigt.

Kapitel 4: Experimenteller Teil 2 - Kollaborative Variation mentaler Modelle

In diesem Kapitel habe ich in drei Experimenten die Leistung von Dyaden beim Unbestimmtheitseffekt getestet. Dazu habe ich im Vergleich zu den Aufgaben aus Kapitel 3 komplexere zweidimensionale Probleme verwendet. In Experiment 3 und 4 habe ich die Leistungen von Dyaden und Individuen als getrennte Stichproben² miteinander verglichen. In Experiment 5 habe ich die Vpn erst individuell und dann in der Dyade oder erst in der Dyade und anschließend individuell getestet. So konnte ich zusätzlich prüfen, wie die Schlussfolgerungen in Abhängigkeit der individuellen Leistungen abgebildet werden. Allgemein ging ich bei allen drei Experimenten davon aus, dass Dyaden mehr valide Konklusionen ziehen als Individuen. Diese Annahme beruht einerseits darauf, dass das geteilte mentale Modell seltener Fehler enthält als individuelle mentale Modelle. Andererseits habe ich angenommen, dass die Dyade in der Lage ist, die Unbestimmtheit besser als Individuen zu erkennen. Das bedeutet, dass die Dyade häufiger die wichtige Annotation über mehrere mögliche Interpretationen verarbeitet und deshalb besonders bei den schwierigen indeterminierten Problemen im Vergleich zu Individuen profitieren sollte. Außerdem habe ich angenommen, dass Personen in einer Dyade in der Lage sein sollten, sich gegenseitig von dem korrekten mentalen Modell zu überzeugen und dieses in ein korrektes geteiltes mentales Modell zu überführen. Deshalb habe ich getestet, ob die Leistung der Dyade am besten durch das *truth wins* Schema nach Davis (1973) abgebildet wird. Dieses Modell habe ich gegen die anderen beiden möglichen Erklärungen für die Gruppenleistung (*equiprobability* und *error wins*) getestet.

Zusammenfassend die drei Hypothesen dieses Kapitels:

Hypothese 1: Dyaden ziehen mehr valide Konklusionen als Individuen.

Hypothese 2: Dyaden ziehen besonders bei indeterminierten Problemen mehr valide Konklusionen als Individuen.

Hypothese 3: Die Leistung der Dyade wird bei indeterminierten Problemen am besten durch das *truth wins* Schema vorhergesagt.

² Die Stichprobe für Experiment 3 habe ich während eines Forschungsaufenthaltes an der Central European University in Budapest (Ungarn) erhoben. Deshalb war dieses Experiment in englischer Sprache. Für die Darstellung des Experimentes in dieser Arbeit habe ich jedoch deutsche Übersetzungen der Prämissen verwendet.

Alle Experimente dieses Kapitels wurden mit OpenSesame (Version 3.1, Mathôt et al., 2012) programmiert und durchgeführt. Die zufällige Zuordnung der Probanden habe ich mit randomizer.org (Version 4, Urbaniak & Plous, 2013) realisiert. Die Aufbereitung der Daten habe ich mit Microsoft Excel (Version 2016) durchgeführt und anschließend mit der Statistiksoftware SPSS (Version 24, IBM Corp., 2016) inferenzstatistisch ausgewertet. Für die Auswertung der relativen Häufigkeiten habe ich die angular transformierten Werte verwendet.

Die vorliegenden Experimente sind im Einklang mit den Standards der aktuellsten Version der Deklaration von Helsinki (World Medical Association, 2013) und wurden von einer lokalen Ethikkommission geprüft. Eine Ausnahme ist dabei, dass für die vorliegenden Experimente keine Pre-Registrierung nach §35 der Deklaration durchgeführt wurde.

4.1 Experiment 3: Kollaborative Variation mit realen Objekten

Experiment 3 ist das erste von drei Experimenten, bei denen ich den Unterschied zwischen Individuen und Dyaden beim Unbestimmtheitseffekt getestet habe. Dazu durften die Vpn reale Objekte auf einem Tisch platzieren und ihre Antworten aus den gelegten Modellen ablesen.

4.1.1 Methoden

4.1.1.1 Stichprobe

Die Stichprobe wurde mit der Teilnehmer Software von Sona Systems (2016) akquiriert. In diesem System können sich die Probanden selbstständig für einen Termin eintragen und werden zufällig einer Bedingung zugeteilt. Da das Experiment in Ungarn und auf Englisch durchgeführt wurde, habe ich für die Untersuchung nur Personen mit einem Englischlevel von mindestens B2 des CEFR berücksichtigt. Weil das Aufgabenmaterial verschiedenfarbige reale Objekte enthielt, mussten zusätzlich alle Personen mit einer Achromasie ausgeschlossen werden. An der Untersuchung nahmen $N = 66$ Vpn (47 weiblich) teil. Das mittlere Alter lag bei $M = 24.8$ ($SD = 4.7$) Jahren und reichte von $MIN = 19$ bis $MAX = 43$ Jahren. Von der Stichprobe arbeiteten $n = 44$ zu zweit (22 Dyaden), die anderen $n = 22$ Personen alleine (Individuum). Die Zuordnung zu der Bedingung war randomisiert. Für den Arbeitsaufwand erhielten die Teilnehmer Wertgutscheine für einen lokalen Supermarkt.

4.1.1.2 Material

Die Vorlage für das Aufgabenmaterial waren die Probleme aus den Studien von Byrne und Johnson-Laird (1989) und Roberts (2000). Dabei bestand dieses Experiment aus zwölf

zweidimensionalen fünf-Term Problemen. Von den zwölf Problemen waren jeweils vier Probleme ein-Modell Probleme, zwei-Modelle determinierte Probleme oder zwei-Modelle indeterminierte Probleme. In Tabelle 7 ist für jede Art der Probleme ein Beispiel abgebildet. Den fünf Termen wurden zufällig verschiedenfarbige (rot, gelb, grün, blau, orange) Legoblöcke zugeordnet, wobei jeder Proband die gleichen Probleme bearbeitete. Da das Experiment mit realen Objekten umgesetzt wurde, habe ich im Gegensatz zu den Originalstudien anstelle von „über“ und „unter“ die Prädikate „vor“ und „hinter“ verwendet. Die vier Prämissen führten immer zu jeweils einer von vier möglichen Rotationen (vgl. Abbildung 4) und es wurde immer die Relation von den zwei herausstehenden Objekten abgefragt. Somit konnten fünf verschiedene Konklusionen aus den Modellen entstehen: „links, rechts, vor, hinter oder keine valide Konklusion“. Das gesamte Aufgabenmaterial mit den prädikatenlogischen Notationen ist Anhang A zu entnehmen. Um sicherzustellen, dass alle Probleme innerhalb einer Kategorie kognitiv ähnlich schwierig sind, habe ich die Onlineversion der von Ragni und Knauff (2013) implementierten Simulation PRISM (*preferred inferences in reasoning with spatial mental models*) verwendet. Dieses Programm gibt mit Hilfe der Prämissen aus, wie viele Operationen für die Lösung eines Problems notwendig sind. Dabei habe ich sichergestellt, dass für die Lösung der Probleme innerhalb einer Kategorie (ein-Modell, zwei-Modelle determiniert und zwei-Modelle indeterminiert) vergleichbar viele Operationen zur Lösung notwendig sind.

Tabelle 7

Beispiele für die im Experiment 3 verwendeten Probleme

	Problem														
	ein-Modell			zwei-Modelle determiniert				zwei-Modelle indeterminiert							
1. Prämisse	Der rote Block ist links von dem gelben Block.			Der rote Block ist links von dem gelben Block.				Der rote Block ist links von dem gelben Block.							
2. Prämisse	Der grüne Block ist rechts von dem gelben Block.			Der grüne Block ist rechts von dem roten Block.				Der grüne Block ist links von dem gelben Block.							
3. Prämisse	Der blaue Block ist vor dem roten Block.			Der blaue Block ist vor dem roten Block.				Der blaue Block ist vor dem roten Block.							
4. Prämisse	Der orangene Block ist vor dem grünen Block.			Der orange Block ist vor dem grünen Block				Der orangene Block ist vor dem grünen Block.							
Modelle	Rot	Gelb	Grün	Rot	Grün	Gelb	Rot	Gelb	Grün	Grün	Rot	Gelb	Rot	Grün	Gelb
	Blau		Orange	Blau	Orange		Blau	Orange		Orange	Blau		Blau	Orange	
Frage	Der blaue Block ist ... von dem orangenen Block.														
Konklusion	Links			Links				Keine valide Konklusion							

Anmerkung: Die Prämissen waren im Original in englischer Sprache und wurden für die Darstellung ins Deutsche übersetzt.



Abbildung 4. Die vier möglichen Rotationen, die für die Probleme verwendet wurden. Dabei wurde immer nach der Relation zwischen den abstehenden (Blau und Orange) Objekten gefragt.

4.1.1.3 Ablauf

Vor Beginn des Experiments wurden die Vpn über die Datenschutzrichtlinien informiert und sie haben eine Einverständniserklärung unterschrieben. Außerdem wurde jeder Proband mit einem Standardtest auf Achromasie getestet (vgl. Anhang B). Das gesamte Experiment fand an einem Computer mit einem (Individuen) oder zwei (Dyade) Monitoren statt und wurde mit einer Videokamera aufgezeichnet³. Neben den Monitoren befand sich eine Box mit den fünf verschiedenfarbigen Legoblöcken. In der Mitte des Tisches befand sich eine graue Legoplatte. Bevor das Experiment begann, wurde den Vpn genau erklärt, wie die Relationen (links, rechts, vor, hinter) zu verstehen sind und es wurde ihnen zusätzlich dazu ein Beispielfoto auf den Tisch gelegt.

Das Experiment lief in zwei Phasen ab. In der ersten Phase wurden den Probanden für 80 Sekunden parallel alle vier Prämissen präsentiert. In dieser Phase sollten die Vpn also ein mentales Modell des Problems konstruieren. Nach den 80 Sekunden fuhr das Experiment automatisch fort. In der zweiten Phase wurden die Prämissen ausgeblendet und die Vpn bekamen einen Satz präsentiert (z. B. „Der blaue Block ist ... von dem orangenen Block“). Die Aufgabe der Vpn war es, diesen Satz zu vervollständigen. Dazu konnten sie mit den Tasten eins bis fünf (für die Antworten: links, rechts, vor, hinter, keine valide Konklusion) auf der Tastatur antworten. Dabei wurde die Zuordnung der Relationen auf die Tasten eins bis vier zufällig ausgewählt. Die Antwort „keine gültige Konklusion“ war jedoch immer auf der

³ Die Videoaufzeichnungen waren ursprünglich für eine explorative Analyse der Diskussionen gedacht. Meine Arbeit fokussiert sich jedoch auf validen Konklusionen und Antwortkombinationen, sodass ich diese Aufzeichnungen nicht ausgewertet habe.

Position mit der Taste fünf. Bei ein-Modell und zwei-Modelle determinierten Problemen war die Anzahl bei der die valide Konklusion „links, rechts, über oder unter“ entsprach, balanciert. Während die Probanden die Konklusion präsentiert bekamen, sollten sie die Legoblöcke aus der Box nehmen und auf der grauen Platte platzieren. Dabei konnten sie sich so viel Zeit nehmen, wie sie wollten und durften die Blöcke jeder Zeit neu positionieren. Wenn sie fertig waren, gaben sie die Antwort mit den Tasten 1 bis 5 auf der Tastatur ein. Danach wurden die Legoblöcke zurück in die Box platziert und die Vpn konnten eine Pause machen und selbstständig fortfahren.

Der Ablauf für die Dyaden unterschied sich in nur wenigen Punkten. Den Dyaden wurde gesagt, dass sie über die Dauer des gesamten Experiments zusammenarbeiten und sie gemeinsam die Antworten finden müssen. Außerdem stand während der ersten Phase (Prämissen) ein Sichtschutz zwischen den beiden Personen und die Vpn durften während dieser Phase nicht miteinander sprechen. Wie auch die Individuen durften die Dyaden erst in der zweiten Phase die Legoblöcke legen und die Modelle diskutieren. Dazu standen ihnen – wie den Individuen – ausschließlich die fünf Legoblöcke zur Verfügung. Das bedeutet, dass die Mitglieder der Dyade gemeinsam nur ein Modell konstruieren und darüber diskutieren konnten. Ein weiterer Unterschied bei der Dyade war, dass – nachdem die eine Person der Dyade die Antwort gegeben hat – die zweite Person bestätigen musste, dass sie mit der Antwort einverstanden war. Da die Dyade am Anfang instruiert wurde, dass sie unbedingt gemeinsam auf eine Lösung kommen musste, war diese Frage lediglich dazu gedacht, die Zusammenarbeit zu fördern. Nach dieser Frage wurde ein Pausenfenster eingeblendet, der Experimentalleiter stellte den Sichtschutz wieder zwischen die Personen und die Vpn legten die Legoblöcke zurück in die Box. Wenn die Vpn beide bereit waren, fuhren sie mit dem Experiment fort. Vor dem Hauptexperiment mit den in randomisierter Abfolge präsentierten zwölf Problemen löste jede Vpn drei Beispielaufgaben. Diese lösten die Vpn – je nach Bedingung – entweder alleine oder in der Dyade. Jedes Beispiel stellte jeweils ein ein-Modell, zwei-Modelle determiniertes oder zwei-Modelle indeterminiertes Problem dar. Wenn die Vpn eine Antwort zu dem Problem gaben, zeigte der Experimentalleiter den Vpn, ob sie die korrekte Antwort gegeben hatten und sagte ihnen zusätzlich die korrekte Lösung. Dabei erklärte er den Vpn die richtige Relation und veranschaulichte die möglichen Modelle mit Hilfe der Legoblöcke auf dem Tisch. Das bedeutet, dass die Vpn eindeutig wussten, welche möglichen Probleme sie erwarteten. Nachdem die Vpn alle zwölf Aufgaben gelöst hatten, wurden die Vpn über das Ziel des Experiments aufgeklärt. Danach wurden ihnen die

Wertgutscheine ausgehändigt und das Experiment war beendet. Die Untersuchung dauerte insgesamt eine Stunde.

4.1.1.4 Design

Das Experiment folgt einem 2×3 gemischten Design mit dem Zwischensubjektfaktor „Gruppe“ (Individuum \times Dyade) und dem Innersubjektfaktor „Problem“ (ein-Modell \times zwei-Modelle determiniert \times zwei-Modelle indeterminiert). Abhängige Variable ist die relative Häufigkeit valider Konklusionen. Das experimentelle Design ist als Überblick mit der Anzahl der Probleme in Tabelle 8 dargestellt.

Tabelle 8

Experimentelles Design mit $N = 66$ Versuchspersonen. Abhängige Variable ist die relative Häufigkeit valider Konklusionen

Gruppe (Zwischensubjektfaktor)	Problem (Innersubjektfaktor)		
	ein-Modell	zwei-Modelle determiniert	zwei-Modelle indeterminiert
Individuen	4 Probleme	4 Probleme	4 Probleme
Dyade	4 Probleme	4 Probleme	4 Probleme

4.1.2 Ergebnisse

Um die Anzahl an Beobachtungen gleich zu halten, habe ich in diesem Experiment nicht die $n = 22$ Datensätze der Individuen zu einer statistischen Dyade gebildet. Das bedeutet, dass für die Auswertung $n = 22$ Datensätze der Individuen mit $n = 22$ Datensätzen der Dyaden verglichen wurden.

4.1.2.1 Valide Konklusionen

Die deskriptiven Daten zeigen, dass ein-Modell Probleme mit 77.27 % ($SD = 23.39$ %) validen Konklusionen am leichtesten waren. Zwei-Modelle determinierte Probleme waren mit 70.45 % ($SD = 24.87$ %) validen Konklusionen etwas schwieriger und die zwei-Modelle indeterminierten Probleme waren mit 61.93 % ($SD = 33.45$ %) validen Konklusionen deskriptiv am schwierigsten zu lösen. Ein-Modell Probleme konnten von Individuen und Dyaden gleichermaßen gut gelöst werden (Individuen: $M = 76.14$ %, $SD = 27.25$ %; Dyaden:

$M = 78.41 \%$, $SD = 19.36 \%$). Bei zwei-Modelle determinierten Problemen konnten Dyaden ($M = 73.86 \%$, $SD = 24.97 \%$) etwas mehr valide Konklusionen ziehen als Individuen ($M = 67.05 \%$, $SD = 24.86 \%$). Der größte deskriptive Unterschied in den validen Konklusionen zeigte sich bei indetermierten Problemen. Dyaden konnten mit 70.45% ($SD = 32.40 \%$) mehr valide Konklusionen ziehen als Individuen ($M = 53.41 \%$, $SD = 33.00 \%$). Abbildung 5 stellt die deskriptiven Daten der Individuen und Dyaden zur Veranschaulichung dar.

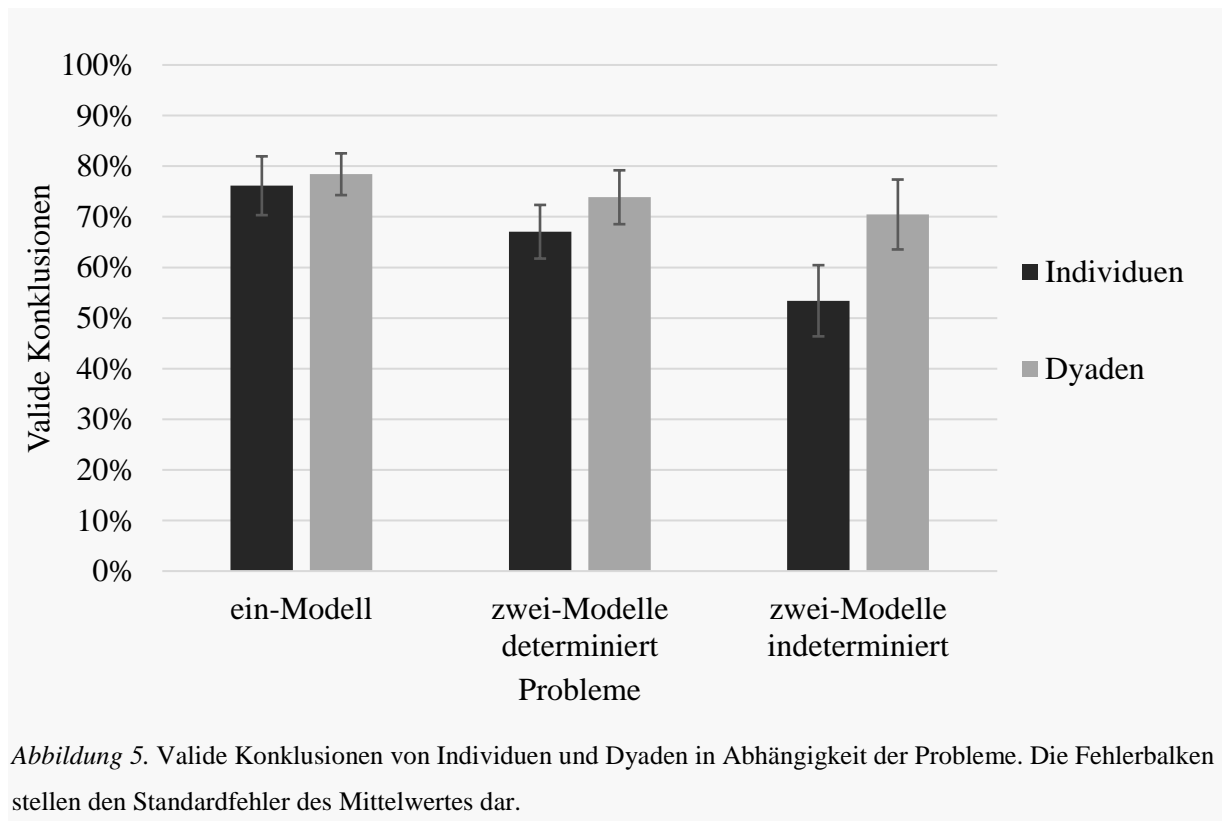


Abbildung 5. Valide Konklusionen von Individuen und Dyaden in Abhängigkeit der Probleme. Die Fehlerbalken stellen den Standardfehler des Mittelwertes dar.

Um die deskriptiven Unterschiede inferenzstatistisch zu testen, habe ich eine 2×3 Varianzanalyse mit Messwiederholung mit dem Zwischensubjektfaktor „Gruppe“ (Individuum \times Dyade) und dem Innersubjektfaktor „Problem“ (ein-Modell \times zwei-Modelle determiniert \times zwei-Modelle indetermiert) durchgeführt. Die Analyse ergab keinen signifikanten Haupteffekt für den Faktor „Gruppe“ ($F(1, 42) = 2.06$, $p = .159$, $\eta_p^2 = .047$) und einen signifikanten Haupteffekt für den Faktor „Problem“ ($F(2, 84) = 3.50$, $p = .035$, $\eta_p^2 = .077$). Die Interaktion Gruppe \times Problem wurde nicht signifikant ($F(2, 84) = 0.83$, $p = .440$, $\eta_p^2 = .019$). Die post-hoc Analyse des Haupteffekts „Problem“ zeigt, dass keiner der drei Einzelvergleiche das nach Bonferroni korrigierte Alpha-Niveau ($\alpha = .017$) unterschreiten konnte. Da die Interaktion keine Signifikanz erreichte, ist es nicht notwendig, weitere Analysen durchzuführen. Weil sich jedoch ein deskriptiver Trend dahingehend zeigte, dass die Leistung der Individuen in Abhängigkeit der Schwierigkeit der Probleme schlechter wurde

(Abbildung 5), habe ich die Leistungen für Individuen und Dyaden jeweils nach einem linearen Trend untersucht. Um für multiples Testen abgesichert zu sein, habe ich für diese Analysen das Alpha-Niveau nach Bonferroni korrigiert ($\alpha = .025$). Es zeigte sich ein signifikanter linearer Trend für Individuen ($F(1, 21) = 6.13, p = .022, \eta_p^2 = .226$), aber kein linearer Trend für Dyaden ($F(1, 21) = 0.72, p = .406, \eta_p^2 = .033$).

4.1.2.2 Entscheidungsschemata

Um die verschiedenen Modelle der Entscheidungsschemata zu testen, habe ich die individuellen relativen Häufigkeiten aus den vier Konklusionen bei indeterminierten Problemen verwendet. Diese relativen Häufigkeiten habe ich mit den drei Modellen (*truth wins*, *equiprobability* und *error wins*) nach Davis (1973) modelliert (siehe auch Stasser, 1999). Dazu werden die individuellen relativen Häufigkeiten aufsteigend sortiert und jeder Wert mit der Entscheidungsmatrix D (vgl. Tabelle 9) multipliziert. Die daraus entstehenden Modelle habe ich gegen die Daten der Dyaden mit dem Kolmogorov-Smirnov Anpassungstest (*goodness of fit*) für zwei Stichproben getestet. Dieser Test prüft, ob sich zwei Verteilungen voneinander unterscheiden. Dazu wird der maximale Abstand – der Wert extremster Differenzen (D_{max}) zwischen diesen Verteilungen – ermittelt. Je höher der D_{max} -Wert, desto schlechter kann das Modell die beobachteten Daten vorhersagen. Da das beste Modell sich so wenig wie möglich von den beobachteten Daten unterscheiden soll, gilt bei diesem Test ein hoher p -Wert ($p > .05$) als wünschenswert. Alle Modelle mit einem $p < .05$ werden als Erklärung für die Daten abgelehnt.

Tabelle 9

Entscheidungsschemata nach Davis (1973) für eine Dyade

r_s, r_n	truth wins		error wins		equiprobability	
	C	I	C	I	C	I
2,0	1	0	1	0	1	0
1,1	1	0	0	1	0.5	0.5
0,2	0	1	0	1	0	1

Anmerkung: r_s = Anzahl der Mitglieder, welche die korrekte Lösung kennen; r_n = Anzahl der Mitglieder, welche die korrekte Lösung nicht kennen; C = Wahrscheinlichkeit, dass die Gruppe die korrekte Lösung nennt; I = Wahrscheinlichkeit, dass die Gruppe die falsche Lösung nennt.

Die durch die Leistung der Individuen vorhergesagten Modelle und die beobachtete Leistung der Dyade sind in Abbildung 6 abgetragen. Man erkennt, dass die Leistung von

Individuen (x-Achse) und Dyaden von $MIN = 0\%$ bis $MAX = 100\%$ validen Konklusionen reicht. Deskriptiv zeigt sich, dass die Leistung der Dyaden in über 50 % der Fälle auf oder über der vom *truth wins* Schema vorhergesagten Leistung liegt. Der Kolmogorov-Smirnov Test zeigt, dass das *truth wins* Schema die beobachtete Leistung der Dyade am besten beschreiben konnte ($D_{max} = .227$, $Z(44) = .75$, $p = .621$). Das *equiprobability* Schema konnte die Daten am nächstbesten abbilden ($D_{max} = .273$, $Z(44) = .91$, $p = .387$) und das *error wins* Schema musste verworfen werden ($D_{max} = .545$, $Z(44) = 1.81$, $p = .003$).

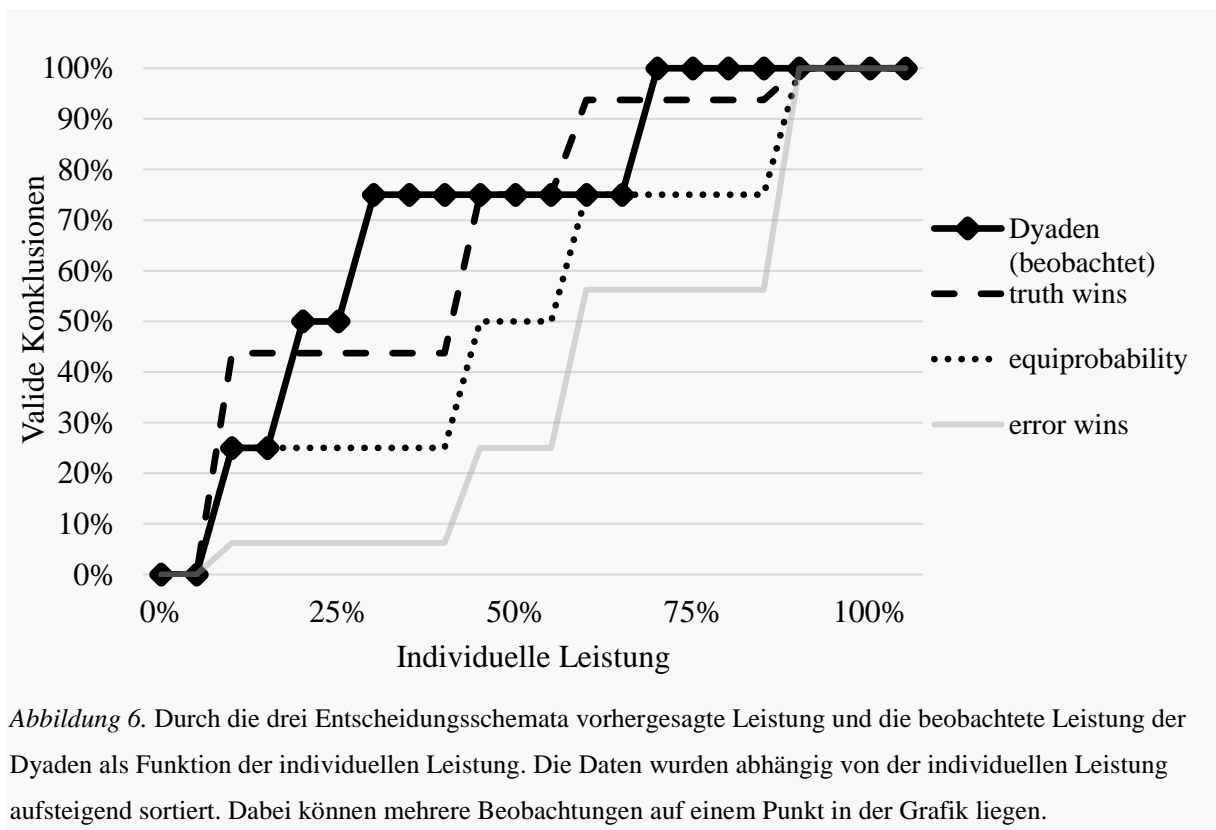


Abbildung 6. Durch die drei Entscheidungsschemata vorhergesagte Leistung und die beobachtete Leistung der Dyaden als Funktion der individuellen Leistung. Die Daten wurden abhängig von der individuellen Leistung aufsteigend sortiert. Dabei können mehrere Beobachtungen auf einem Punkt in der Grafik liegen.

4.1.3 Diskussion

Dyaden konnten grundsätzlich nicht mehr valide Konklusionen ziehen als Individuen. Deshalb sprechen meine Daten gegen die erste Hypothese (*Dyaden ziehen mehr valide Konklusionen als Individuen*). Die deskriptiven Daten zeigen uns jedoch, dass Dyaden gerade bei schwierigen Problemen mehr valide Konklusionen ziehen konnten. So erreichten sie bei indeterminierten Problemen 70 % valide Konklusionen. Individuen kamen hingegen nur auf 53 % valide Konklusionen. Im direkten Vergleich waren Dyaden zwar nicht signifikant besser, aber Individuen zeigten einen klaren linearen Trend: Umso schwieriger das Problem, desto mehr Fehler entstanden bei den Individuen. Dyaden hingegen zeigten diesen Trend nicht. Das bedeutet, dass sie unabhängig vom Problem (ein-Modell, zwei-Modelle

determiniert, zwei-Modelle indeterminiert) eine ähnliche Leistung zeigen konnten. Meine Daten geben also einen Hinweis darauf, dass Dyaden besonders bei schwierigen Problemen profitieren. Deshalb kann dieses Experiment meine zweite Hypothese, dass *Dyaden besonders bei indeterminierten Problemen mehr valide Konklusionen ziehen*, teilweise unterstützen. Diese Annahme wird auch durch die Art der Fehler bei den indeterminierten Problemen gestützt. Bei indeterminierten Problemen müssen für eine korrekte Antwort immer zwei Modelle berücksichtigt werden. Fehler können also darauf zurückzuführen sein, dass eine Vpn zumindest ein korrektes Modell konstruiert hat. In diesem Fall würden sie anstelle von „keiner validen Konklusion“ mit einer Relation (z. B. links) antworten. Wenn man die korrekten Antworten und alle Fehler, bei denen die Antwort der Vpn auf zumindest ein korrektes Modell zurückzuführen sind, miteinander addiert, erklären diese Möglichkeiten 100 % der Antworten von den Dyaden. Bei den Individuen ist dies anders. Individuen machten häufiger Fehler, die nicht auf ein korrektes Modell zurückzuführen sind. Wenn man die korrekten Antworten und jene, die auf zumindest ein korrektes Modell zurückzuführen sind, addiert, können nur 92 % dieser Antworten erklärt werden. Die anderen acht Prozent sind auf fehlerhafte Modelle zurückzuführen. Dieser Unterschied wirkt nicht sehr groß, zeigt aber auf der einen Seite, dass Dyaden in der Lage sind, ein kohärentes geteiltes mentales Modell zu konstruieren, welches den individuellen Modellen überlegen ist. Auf der anderen Seite könnte dies ein Hinweis dafür sein, dass die Fehler der Dyade dadurch entstehen, dass sich die Gruppe auf das präferierte mentale Modell einer Person einigt und die Annotation über weitere Interpretationen der Prämissen nicht beachtet.

Dass aber die Dyade sehr häufig in der Lage ist, eine vollständige geteilte Repräsentation der Probleme zu konstruieren, zeigen die Befunde der Entscheidungsschemata. Die Gruppenmitglieder waren in der Lage, sich gegenseitig von der korrekten Repräsentation zu überzeugen. Deshalb konnte das *truth wins* Schema die Leistung der Dyade am besten vorhersagen. Dieses Ergebnis stützt meine dritte Hypothese (*Die Leistung der Dyade bei indeterminierten Problemen wird am besten durch das truth wins Schema vorhergesagt*). Da das *error wins* Schema verworfen werden musste, nehme ich an, dass die Dyade keiner starken Akzentuierung ihrer kognitiven Prozesse beim Denken mit räumlichen Relationen unterliegt. Würden beide Individuen einer Dyade lediglich ein präferiertes mentales Modell konstruieren und sich gegenseitig von diesem Modell überzeugen, ohne dieses Modell zu variieren, hätte das *error wins* Schema die Leistung der Dyaden am besten vorhersagen müssen. Außerdem ist anzumerken, dass das *equiprobability* Schema nicht ganz verworfen werden konnte. Dieses Entscheidungsschema besagt, dass die

Personen in der Dyade sich auf keine eindeutige Lösung einigen können und deshalb zufällig eine Antwort wählen. Da die Dyade jedoch häufig nicht nur die vom *equiprobability* Schema, sondern sogar die Leistung nach dem *truth wins* Schema übertreffen konnte, ist es sehr unwahrscheinlich, dass das *equiprobability* Schema die Leistung der Dyade erklären kann. Würden die Personen innerhalb der Dyade lediglich zufällig eine Konklusion ziehen, müssten insgesamt die Leistungen der Dyade deutlich niedriger sein. Auch wenn das *truth wins* Schema die Leistung der Dyade am besten beschreiben konnte, ist es aber wichtig, diese Befunde durch weitere Experimente zu testen, um eine klarere Abgrenzung zum *equiprobability* Schema nachzuweisen.

Zusammenfassend kann mein Experiment zeigen, dass Dyaden das Potential haben, die individuelle Leistung zu übertreffen. Obwohl sich Personen in einer Dyade gegenseitig von den korrekten mentalen Modellen überzeugen können und Dyaden durch schwierigere Probleme kaum beeinflusst wurden, können meine Ergebnisse nicht eindeutig sagen, dass Dyaden besonders bei den schwierigen indeterminierten Problemen profitieren. Außerdem muss an dem vorliegenden Experiment kritisiert werden, dass ich nicht sicher sagen kann, dass die Leistungen – von Individuen und Dyaden – wirklich auf schlussfolgernde Denkprozesse zurückzuführen sind. Da die Probanden reale Objekte verwenden konnten, mussten sie sich lediglich die Anordnung merken und keine mentalen Informationen aus dem Modell ablesen. Da die Vpn die Objekte auf den Tisch platzieren mussten, konnten sie die Konklusion einfach aus der Anordnung der Legoblöcke ablesen. Um dieser Kritik entgegenzuwirken, habe ich das Experiment 4 ohne reale Objekte durchgeführt.

4.2 Experiment 4: Kollaborative Variation ohne reale Objekte

Experiment 4 ist das zweite Experiment von den drei Experimenten zur kollaborativen Variation mentaler Modelle. In diesem Experiment habe ich auf reale Objekte verzichtet. Dadurch sollten die Leistungen eindeutiger auf kognitive Prozesse und nicht auf das Ablesen einer Anordnung auf dem Tisch zurückzuführen sein.

4.2.1 Methoden

4.2.1.1 Stichprobe

Die Stichprobe wurde mit Hilfe des Rundmail-Systems der Justus-Liebig-Universität akquiriert. Dabei wurden keine expliziten Ausschlusskriterien vordefiniert. An der Untersuchung nahmen $N = 84$ Vpn (63 weiblich) teil. Das mittlere Alter lag bei $M = 23.3$ ($SD = 4.5$) Jahren und reichte von $MIN = 18$ bis $MAX = 52$ Jahre. Von der Stichprobe arbeiteten n

= 42 zu zweit (21 Dyaden) und die anderen $n = 42$ Personen alleine (Individuum). Die Zuordnung zu der Bedingung war dabei randomisiert. Für den Arbeitsaufwand erhielten die Vpn Versuchspersonenstunden oder 8€ je Stunde.

4.2.1.2 Material

Für dieses Experiment wurden die Probleme aus Experiment 3 verwendet. Das Aufgabenmaterial wurde jedoch ins Deutsche übersetzt. Außerdem durften die Probanden keine realen Objekte zum Legen ihrer mentalen Modelle verwenden. Deshalb wurden die Terme von verschiedenfarbigen (rot, gelb, grün, blau und orange) Blöcken in „Quadrate“ umbenannt. Zusätzlich habe ich die Prädikate „vor“ und „hinter“ durch „unter“ und „über“ ersetzt. Dies war notwendig, da die (Beispiel-)Modelle nicht mehr im dreidimensionalen Raum auf der Lego Platte, sondern auf dem zweidimensionalen PC Monitor dargestellt wurden. Insgesamt ergeben sich also wieder zwölf Probleme, von denen jeweils vier ein-Modell, zwei-Modelle determinierte und zwei-Modelle indetermierte Probleme waren.

4.2.1.3 Ablauf

Der Ablauf war nahezu identisch zu Experiment 3. Die Probanden wurden zunächst über den Datenschutz informiert und unterschrieben eine Einverständniserklärung. Die Instruktionen für das Experiment wurden danach auf einem Computerbildschirm präsentiert. Individuen arbeiteten alleine vor einem Monitor und die Vpn der Dyade saßen jeweils vor einem Monitor. Das Experiment lief wieder in zwei Phasen ab. Während der ersten Phase wurden die Prämissen für 80 Sekunden präsentiert. Danach fuhr das Experiment automatisch fort und die Prämissen wurden ausgeblendet. Bei den Dyaden stand zwischen den beiden Monitoren während der ersten Phase (Prämissen) ein Sichtschutz und die Vpn durften erst miteinander sprechen, wenn die Prämissen nach 80 Sekunden verschwanden und die Frage präsentiert wurde. Die Vpn haben die Antworten wieder über die Tastatur mit den Tasten eins bis fünf gegeben. Fünf entsprach immer „keine valide Konklusion“ und den Tasten eins bis vier wurde zufällig „links, rechts, über, unter“ zugeordnet. Bei ein-Modell und zwei-Modelle determinierten Problemen war die Anzahl, bei der die korrekte Antwort „links, rechts, über oder unter“ entsprach, balanciert. Vor dem Hauptexperiment mit zwölf in randomisierter Abfolge präsentierten Problemen (jeweils vier ein-Modell, zwei-Modelle determiniert und zwei-Modelle indetermierte Probleme) haben die Vpn drei Übungsaufgaben (für jedes der drei möglichen Probleme ein Beispiel) durchgeführt. Dabei hat der Experimentalleiter den Vpn nach jeder Aufgabe genau erklärt, welches die valide Konklusion ist und wie diese Konklusion erklärt werden kann. Nachdem die Vpn alle zwölf Probleme des

Hauptexperiments gelöst hatten, wurden sie über das Ziel des Experiments aufgeklärt. Das Experiment dauerte insgesamt eine Stunde.

4.2.1.4 Design

Das Experiment folgt einem 2×3 gemischten Design mit dem Zwischensubjektfaktor „Gruppe“ (Individuum \times Dyade) und dem Innersubjektfaktor „Problem“ (ein-Modell \times zwei-Modelle determiniert \times zwei-Modelle indeterminiert). Abhängige Variable ist die relative Häufigkeit valider Konklusionen. Das experimentelle Design ist als Überblick mit der Anzahl der Probleme in Tabelle 10 dargestellt.

Tabelle 10

Experimentelles Design mit $N = 84$ Versuchspersonen. Abhängige Variable ist die relative Häufigkeit valider Konklusionen

Gruppe (Zwischensubjektfaktor)	Problem (Innersubjektfaktor)		
	ein-Modell	zwei-Modelle determiniert	zwei-Modelle indeterminiert
Individuen	4 Probleme	4 Probleme	4 Probleme
Dyade	4 Probleme	4 Probleme	4 Probleme

4.2.2 Ergebnisse

Aus den Daten der $n = 42$ Individuen habe ich zufällig jeweils zwei Personen ausgewählt. Aus den Leistungen dieser beiden Personen habe ich dann den Mittelwert ermittelt und so eine „statistische Dyade“ gebildet. Das bedeutet, dass ich die Beobachtungen gleich gehalten habe und für die Auswertung der Daten die relativen Häufigkeiten von $n = 21$ Individuen und $n = 21$ Dyaden verwendet habe.

4.2.2.1 Valide Konklusionen

Die deskriptiven Daten zeigen, dass ein-Modell Probleme mit 90.18 % ($SD = 13.95$ %) validen Konklusionen am leichtesten waren. Zwei-Modelle determinierte Probleme waren mit 85.12 % ($SD = 19.57$ %) validen Konklusionen etwas schwieriger und die zwei-Modelle indeterminierten Probleme waren mit 80.95 % ($SD = 17.94$ %) validen Konklusionen deskriptiv am schwierigsten zu lösen. Dyaden konnten deskriptiv bei allen Problemen mehr valide Konklusionen ziehen als Individuen. Bei ein-Modell Problemen erreichten Dyaden

96.43 % ($SD = 08.96$ %) und Individuen 83.93 % ($SD = 15.38$ %) valide Konklusionen. Bei zwei-Modelle determinierten Problemen war der Unterschied am größten und Dyaden konnten 95.24 % ($SD = 10.06$ %) valide Konklusionen ziehen (Individuen: $M = 75.00$ %, $SD = 21.65$ %). Zwei-Modelle indeterminierte Probleme wurden von Dyaden mit 89.29 % ($SD = 12.68$ %) validen Konklusionen und von Individuen mit 72.62 % ($SD = 18.80$ %) validen Konklusionen gelöst. Abbildung 7 stellt die deskriptiven Daten der Individuen und Dyaden zur Veranschaulichung dar.

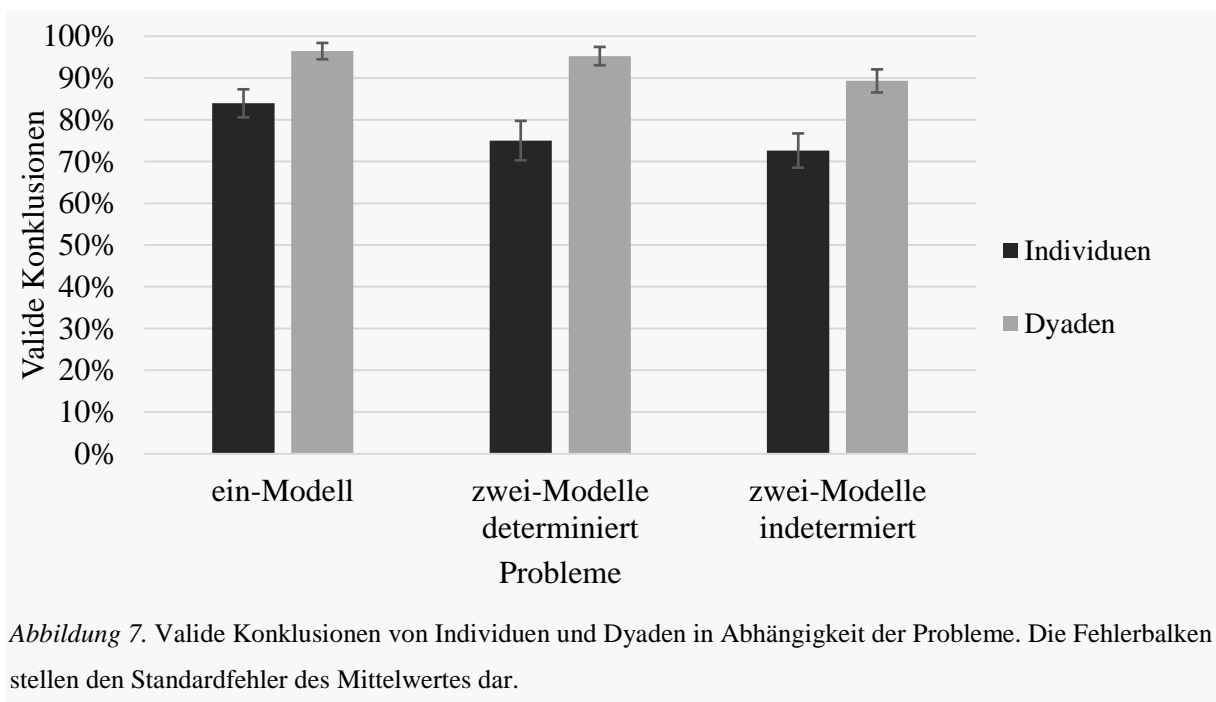


Abbildung 7. Valide Konklusionen von Individuen und Dyaden in Abhängigkeit der Probleme. Die Fehlerbalken stellen den Standardfehler des Mittelwertes dar.

Um die deskriptiven Unterschiede inferenzstatistisch zu testen, habe ich eine 2×3 Varianzanalyse mit Messwiederholung mit dem Zwischensubjektfaktor „Gruppe“ (Individuum \times Dyaden) und dem Innersubjektfaktor „Problem“ (ein-Modell \times zwei-Modelle determiniert \times zwei-Modelle indeterminiert) durchgeführt. Die Analyse ergab einen signifikanten Haupteffekt für den Faktor „Gruppe“ ($F(1, 40) = 29.22, p < .001, \eta_p^2 = .422$). Dyaden konnten mehr valide Konklusionen ziehen als Individuen. Weiterhin zeigte sich ein signifikanter Haupteffekt für den Faktor „Problem“ ($F(2, 80) = 4.93, p = .010, \eta_p^2 = .110$). Nur einer der post-hoc Tests wurde unter dem nach Bonferroni korrigierten Alpha-Niveau ($\alpha = .017$) signifikant. Ein-Modell Probleme waren signifikant leichter als zwei-Modelle indeterminierte Probleme ($t(41) = 2.96, p = .005, d = .57$). Die Interaktion Gruppe \times Problem wurde nicht signifikant ($F(2, 80) = 0.72, p = .489, \eta_p^2 = .018$).

4.2.2.2 Entscheidungsschemata

Um die verschiedenen Modelle der Entscheidungsschemata (*truth wins*, *equiprobability*, *error wins*) zu testen, habe ich die gemittelten validen Konklusionen von den $n = 21$ Individuen bei indeterminierten Problemen verwendet und die vorhergesagten Schemata mit der Leistung der Dyaden verglichen. Die durch die Leistung der Individuen vorhergesagten Modelle und die beobachtete Leistung der Dyade sind in Abbildung 8 abgetragen. Man erkennt, dass die Leistung von Individuen (x-Achse) von $MIN = 38\%$ bis $MAX = 100\%$ validen Konklusionen reicht. Die Leistung der Dyaden reicht hingegen von $MIN = 75\%$ bis $MAX = 100\%$ validen Konklusionen. Deskriptiv zeigt sich außerdem, dass die Leistung der Dyaden in ungefähr der Hälfte der Fälle über der von dem *truth wins* Schema vorhergesagten Leistung liegt. Der Kolmogorov-Smirnov Test zeigt, dass das *truth wins* Schema die beobachtete Leistung der Dyade am besten beschreibt ($D_{max} = .381$, $Z(42) = 1.23$, $p = .095$). Das *equiprobability* Schema ($D_{max} = .476$, $Z(42) = 1.54$, $p = .017$) musste genau wie das *error wins* Schema verworfen werden ($D_{max} = .667$, $Z(42) = 2.16$, $p < .001$).

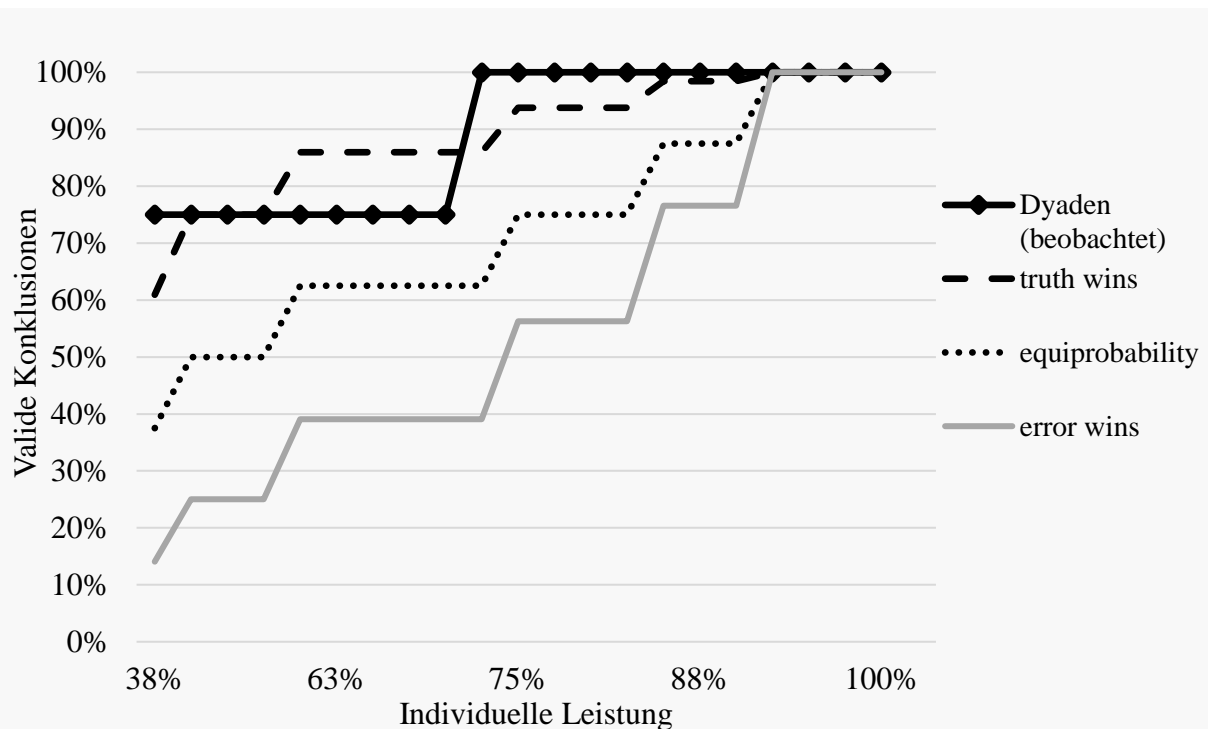


Abbildung 8. Durch die drei Entscheidungsschemata vorhergesagte Leistung und die beobachtete Leistung der Dyaden als Funktion der individuellen Leistung. Die Daten wurden abhängig von der individuellen Leistung aufsteigend sortiert. Dabei können mehrere Beobachtungen auf einem Punkt in der Grafik liegen.

4.2.3 Diskussion

In diesem Experiment nutzten die Vpn keine realen Objekte, um die Konklusionen zu ziehen. Das Denken von Individuen fand ausschließlich auf mentaler Ebene statt und Dyaden konnten ihre Informationen miteinander teilen, indem sie über die möglichen Modelle sprachen und diese diskutierten. Das bedeutet, dass die Lösung dieser Aufgaben deutlich mehr auf epistemische Denkprozesse zurückzuführen ist. Die Ergebnisse zeigen, dass Dyaden insgesamt 16 % mehr valide Konklusionen erreichen als Individuen. Die Daten stützen also meine erste Hypothese, dass *Dyaden mehr valide Konklusionen ziehen als Individuen*. Dyaden sind in der Lage, ihre zunächst individuell konstruierten Modelle miteinander zu teilen und in ein kohärentes geteiltes mentales Modell zu integrieren. Das bedeutet, dass sie – wie in der Literatur zum Problemlösen und Denken in Gruppen (Laughlin et al., 2002, 2003; Moshman & Geil, 1998; Maciejovsky & Budescu, 2007) – mehr Informationen verarbeiten und mehr valide Konklusionen ziehen können. Obwohl der Unterschied bei schwierigeren Problemen deskriptiv größer wurde, konnten Dyaden bei den schwierigsten indeterminierten Problemen nicht besonders profitieren. Deshalb sprechen die Ergebnisse von diesem Experiment gegen meine zweite Hypothese, dass *Dyaden besonders bei indeterminierten Problemen mehr valide Konklusionen ziehen als Individuen*. Dieser Befund könnte durch zwei Faktoren erklärt

werden. Zum einen wurde allen Vpn zu Beginn des Experimentes ausführlich erläutert, wie die verschiedenen Aufgaben zu lösen sind. Das bedeutet, dass sie instruiert wurden, bei indeterminierten Problemen ihr mentales Modell zu variieren. Dass die Individuen diese Instruktion berücksichtigen konnten, zeigt sich in ihren sehr guten Leistungen der Individuen bei indeterminierten Problemen (72 %). Dabei unterscheiden sich die Leistungen ebenfalls nur geringfügig von zwei-Modell determinierten Problemen (75 %). Dieser Befund konnte in anderen Untersuchungen zum Unbestimmtheitseffekt nicht nachgewiesen werden (z. B. Byrne & Johnson-Laird, 1989; Ragni & Knauff, 2013). Wenn die Vpn nicht explizit über die Unbestimmtheit instruiert werden, ist es durchaus denkbar, dass die Aufgaben insgesamt schwieriger werden. In diesem Fall könnten Dyaden mehr von der Interaktion profitieren und sich gegenseitig häufiger auf die Annotation der Unbestimmtheit aufmerksam machen. Eine weitere Erklärung dafür, dass die Dyaden bei indeterminierten Problemen nicht am meisten profitieren konnten ist, dass sie nicht von Beginn an über die Modelle und Prämissen diskutieren konnten. Dieses Experiment war so konzipiert, dass jede Vpn in einer Dyade zunächst ein individuelles Modell konstruiert und erst dann in der Diskussionsphase diese beiden Modelle in ein geteiltes mentales Modell überführt. Es könnte also sein, dass diese Methodik dazu führte, dass die Vpn in der Dyade die Annotation über eine Unbestimmtheit der Probleme während der ersten Phase (Prämissen) bereits vergessen hatten. Somit hätten sie die Annotation in der Diskussionsphase gar nicht miteinander teilen können. Diese Annahme lässt sich – wie schon in Experiment 3 – durch die Fehler der Vpn stützen. Alle Fehler der Dyaden waren auf mindestens ein korrektes Modell zurückzuführen. Bei Individuen hingegen konnten die Fehler nicht immer auf ein korrektes Modell zurückgeführt werden. Allgemein waren in diesem Experiment die Leistungen von Individuen und Dyaden bei indeterminierten Problemen sehr hoch. Dies zeigt sich auch in der Analyse zu den Entscheidungsschemata. Die Leistung von Individuen reicht von 38 bis 100 % validen Konklusionen. Dyaden waren so gut, dass ihre schlechteste Leistung bei 75 % und ihre beste Leistung bei 100 % lag. Diese Leistung der Dyade konnte deskriptiv sogar die vom *truth wins* Modell vorhergesagte Leistung übertreffen. Dennoch war das *truth wins* Modell das beste von den drei Modellen, um die Leistung der Dyade vorherzusagen. Deshalb stützen die Befunde meine Hypothese, dass *die Leistung der Dyade bei indeterminierten Problemen am besten durch das truth wins Schema vorhergesagt wird.*

Das vorliegende Experiment konnte also zeigen, dass Dyaden mehr valide Konklusionen ziehen als Individuen. Außerdem konnte die Leistung wieder am besten durch das *truth wins* Modell vorhergesagt werden. Das bedeutet, dass die Personen in einer Dyade

sich gegenseitig von den korrekten Repräsentationen überzeugen konnten. Obwohl die Dyade bei indeterminierten Problemen die vorhergesagte Leistung des *truth wins* Modells übertreffen konnte, zeigte sich kein klarer Vorteil der Gruppe für die besonders schwierigen indeterminierten Probleme. Dies könnte durch einen von der Instruktion hergeleiteten Deckeneffekt erklärt werden. Deshalb habe ich das nachfolgende Experiment mit einer anderen Instruktion durchgeführt.

4.3 Experiment 5: Kollaborative Variation und Reihenfolgeeffekte

In diesem Experiment habe ich den Vpn vorher nicht explizit die Lösungen der verschiedenen Probleme gezeigt. Dies sollte die Lösung der Probleme allgemein schwieriger machen und somit zu größeren Unterschieden zwischen Individuen und Dyaden führen. Außerdem habe ich die Vpn, wie bei den von mir im Kapitel 1.4.3.2 vorgestellten Experimenten von Moshman und Geil (1998) und Maciejovsky und Budescu (2007) in verschiedenen Reihenfolgen untersucht. Die Vpn haben die Probleme erst alleine und dann ein weiteres Mal in der Dyade gelöst. Mit dieser Methode konnte ich nicht nur die theoretischen Entscheidungsschemata wie das *truth wins* testen. Vielmehr konnte ich durch das vorherige individuelle Lösen der Aufgaben die Verteilung von korrekten individuellen Leistungen und deren spätere Gruppenleistung vergleichen und ein eigenes Modell für das kollaborative räumliche Schlussfolgern testen.

4.3.1 Methoden

4.3.1.1 Stichprobe

Die Stichprobe wurde mit Hilfe des Rundmail-Systems der Justus-Liebig-Universität akquiriert. Dabei wurden keine expliziten Ausschlusskriterien vordefiniert. An der Untersuchung nahmen $N = 70$ Vpn⁴ (53 weiblich) teil. Das mittlere Alter lag bei $M = 23.3$ ($SD = 3.4$) Jahren und reichte von $MIN = 18$ bis $MAX = 35$ Jahre. Von der Stichprobe arbeiteten $n = 36$ zunächst alleine und dann in der Dyade (Individuum/Dyade). Die anderen $n = 34$ Personen arbeiteten erst in der Dyade und daraufhin alleine (Dyade/Individuum). Die Zuordnung zu der Bedingung war dabei randomisiert. Für den Arbeitsaufwand erhielten die Vpn Versuchspersonenstunden oder 8€ pro Stunde.

⁴ Bei $n = 6$ Vpn konnten die demografischen Daten durch einen technischen Fehler nicht erhoben werden. Deshalb beziehen sich die demografischen Daten auf $n = 63$ Vpn. Die Datensätze des Hauptexperimentes enthalten jedoch alle $N = 70$ Vpn

4.3.1.2 Material

Das Material bestand wieder aus den zwölf Aufgaben, die ich bereits in Experiment 3 und 4 verwendet habe (vgl. Anhang A für die Prädikatenlogischen Schemata.). Dabei waren vier Probleme jeweils ein-Modell, zwei-Modelle determinierte oder zwei-Modelle indeterminierte Probleme. Für die Terme habe ich Objekte aus den Kategorien „Obst“ (Apfel, Birne, Mango, Kiwi, Orange) oder „Werkzeug“ (Hammer, Säge, Meißel, Bohrer, Feile) verwendet. Dies sollte einen realistischeren Kontext schaffen und die Verarbeitung der Probleme erleichtern. Zudem konnten somit ausreichend unterschiedliche Probleme für die zwei verschiedenen experimentellen Phasen (Individuen vs. Dyade) konstruiert werden.

4.3.1.3 Ablauf

Das Experiment lief ähnlich zu den anderen Experimenten ab. Eine wichtige Änderung war jedoch, dass es zwei verschiedene Reihenfolgen gab. In der einen Reihenfolge bearbeiteten die Vpn zunächst alle zwölf Probleme alleine und darauf ein weiteres Mal mit einem Partner (Individuum/Dyade). In der anderen Reihenfolge war dies genau umgekehrt. Die Vpn lösten die Probleme zunächst in der Dyade und danach ein weiteres Mal alleine (Dyade/Individuum). Diese Reihenfolge habe ich verwendet, um zu prüfen, ob Unterschiede nicht ausschließlich auf Lerneffekte zurückzuführen sind. Wenn die Vpn alleine arbeiteten, saßen sie in schallisolierten Versuchskabinen. Wenn die Vpn die Probleme gemeinsam in einer Dyade lösten, saßen sie zusammen vor einem 32“ Monitor. Damit niemand in den beiden Durchgängen die identischen Probleme zweimal lösen musste, wurden die Terme für den zweiten Durchgang geändert. Wenn die Personen im ersten Durchgang die Probleme mit Termen aus der Kategorie „Obst“ bearbeiteten, haben sie im zweiten Durchgang Terme aus der Kategorie „Werkzeug“ (oder umgekehrt) präsentiert bekommen. Insgesamt bearbeitete jede Vpn also 24 Aufgaben. Zwölf von diesen Problemen bearbeitete sie alleine und zwölf Probleme in der Dyade. In beiden Durchgängen wurden dabei die Probleme in randomisierter Abfolge präsentiert.

Zu Beginn des Experimentes wurden die Vpn über die Datenschutzrichtlinien aufgeklärt und sie unterschrieben eine Einverständniserklärung. Die Instruktionen wurden den Vpn auf einem PC Monitor präsentiert. Dabei wurde ihnen gesagt, dass die Aufgaben manchmal schwieriger seien als sie zunächst erscheinen und dass sie genau überlegen sollen, um die korrekte Lösung zu finden. Den Dyaden wurde zusätzlich gesagt, dass sie gemeinsam eine Lösung der Probleme finden müssen. Alle Aufgaben inklusive der möglichen Konklusionen wurden auf einer Seite präsentiert. Das bedeutet, dass die Informationen

(Prämissen und Frage) parallel zur Verfügung standen und (in der Dyade) besprochen werden konnten. Gemessen wurde die Anzahl der validen Konklusionen. Vor jedem Teil des Hauptexperimentes (Individuum oder Dyade) bearbeiteten alle Vpn drei Übungsaufgaben (jeweils ein ein-Modell, zwei-Modelle determiniertes und zwei-Modelle indeterminiertes Problem). Dabei erhielten die Vpn kein Feedback über die Richtigkeit ihrer Antwort. Außerdem wurden ihnen – im Vergleich zu Experiment 3 und 4 – die richtigen Lösungen der Übungsaufgaben nicht explizit vorgezeigt. Während der Übungsaufgaben konnten die Vpn beliebig viele Fragen stellen, wobei die Lösungen nicht explizit erklärt wurden. Nachdem die Vpn alle 24 Probleme des Hauptexperiments beantwortet hatten, wurden sie über das Ziel des Experiments aufgeklärt. Das gesamte Experiment dauerte circa eineinhalb Stunden.

4.3.1.4 Design

Das Experiment folgt einem $2 \times 2 \times 3$ gemischten Design mit dem Zwischensubjektfaktor „Reihenfolge“ (Individuum/Dyade \times Dyade/Individuum), dem Innersubjektfaktor „Problem“ (ein-Modell \times zwei-Modelle determiniert \times zwei-Modelle indeterminiert) und dem Innersubjektfaktor „Gruppe“ (Individuum \times Dyade). Abhängige Variable ist die relative Häufigkeit valider Konklusionen. Das experimentelle Design ist als Überblick mit der Anzahl der Probleme in Tabelle 11 dargestellt.

Tabelle 11

Experimentelles Design mit $N = 70$ Versuchspersonen. Abhängige Variable ist die relative Häufigkeit valider Konklusionen

Reihenfolge (Zwischensubjekt)	Gruppe (Innersubjekt)	Problem (Innersubjekt)		
		ein-Modell	zwei-Modelle determiniert	zwei-Modelle indeterminiert
Erst alleine	Individuum	4 Probleme	4 Probleme	4 Probleme
	Dyade	4 Probleme	4 Probleme	4 Probleme
Erst Dyade	Dyade	4 Probleme	4 Probleme	4 Probleme
	Individuum	4 Probleme	4 Probleme	4 Probleme

4.3.2 Ergebnisse

Um die Anzahl der Beobachtungen gleich zu halten, habe ich die relativen Häufigkeiten der validen Konklusionen von jeweils den Individuen, die auch gemeinsam in einer Dyade arbeiteten, gemittelt. Das bedeutet, dass insgesamt $n = 35$ Datensätze für Individuen und $n = 35$ Datensätze für Dyaden für die Berechnungen verwendet wurden. Um die deskriptiven Daten des komplexen Designs übersichtlicher darzustellen, werde ich sie nach der „Reihenfolge“ (Individuum/Dyade oder Dyade/Individuum) getrennt berichten. Die inferenzstatistische Analyse erfolgt mit allen Daten zusammen.

4.3.2.1 Valide Konklusionen

Individuum/Dyade: Die deskriptiven Daten zeigen, dass alle Probleme ähnlich schwierig zu lösen waren (ein-Modell Probleme: $M = 72.22\%$, $SD = 25.32\%$; zwei-Modelle determinierte Probleme: $M = 73.40\%$, $SD = 24.91\%$; zwei-Modelle indetermierte Probleme: $M = 73.61\%$, $SD = 21.78\%$). Dyaden konnten deskriptiv bei allen Problemen mehr valide Konklusionen ziehen als Individuen. Bei ein-Modell Problemen erreichten Dyaden 80.56% ($SD = 23.57\%$) und Individuen 63.89% ($SD = 27.08\%$) valide Konklusionen. Bei zwei-Modelle determinierten Problemen haben Dyaden 84.72% ($SD = 28.62\%$) valide Konklusionen gezogen und Individuen 63.19% ($SD = 21.21\%$). Bei zwei-Modelle indetermierten Problemen kamen Dyaden auf 86.11% ($SD = 23.04\%$) valide Konklusionen und Individuen erreichten mit 61.11% ($SD = 20.51\%$) validen Konklusionen die deskriptiv schlechteste Leistung. Abbildung 9 stellt die deskriptiven Daten der Reihenfolge „Individuen/Dyade“ zur Veranschaulichung dar.

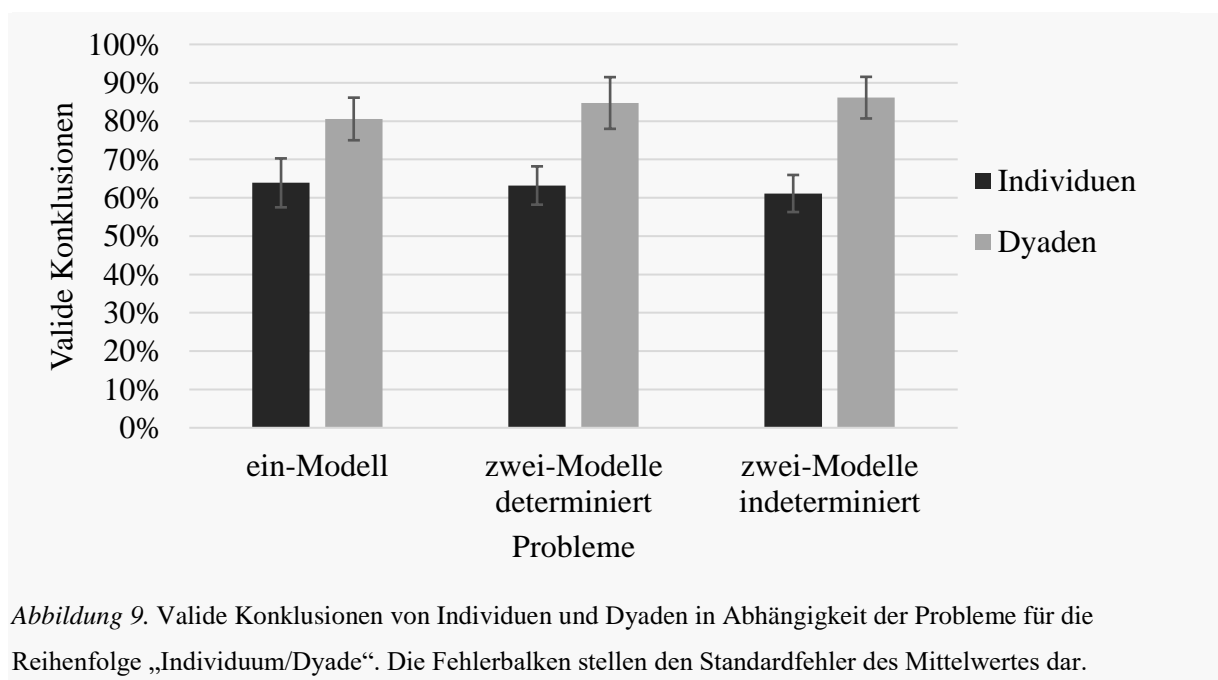


Abbildung 9. Valide Konklusionen von Individuen und Dyaden in Abhängigkeit der Probleme für die Reihenfolge „Individuum/Dyade“. Die Fehlerbalken stellen den Standardfehler des Mittelwertes dar.

Dyade/Individuum: Bei dieser Reihenfolge zeigte sich, dass die deskriptive Leistung bei ein-Modell Problemen mit 77.94 % ($SD = 21.83$ %) vergleichbar mit der Leistung von zwei-Modelle determinierten Problemen ($M = 75.74$ %, $SD = 25.64$ %) war. Zwei-Modelle indetermierte Probleme waren mit 65.44 %, ($SD = 32.85$ %) deskriptiv am schwierigsten. Dyaden konnten bei ein-Modell Problemen mit 82.35 % ($SD = 22.99$ %) mehr valide Konklusionen ziehen als Individuen ($M = 73.53$ %, $SD = 20.67$ %). Bei zwei-Modelle determinierten Problemen erreichten Dyaden mit 75.00 % ($SD = 27.95$ %) eine vergleichbare Leistung wie Individuen ($M = 76.47$ %, $SD = 23.34$ %). Bei zwei-Modelle indetermierten Problemen erreichten Dyaden mit 67.65 % ($SD = 32.79$ %) deskriptiv ebenfalls eine vergleichbare Leistung zu Individuen ($M = 63.24$ %, $SD = 32.92$ %). Abbildung 10 stellt die deskriptiven Daten der Reihenfolge „Dyade/Individuum“ zur Veranschaulichung dar.

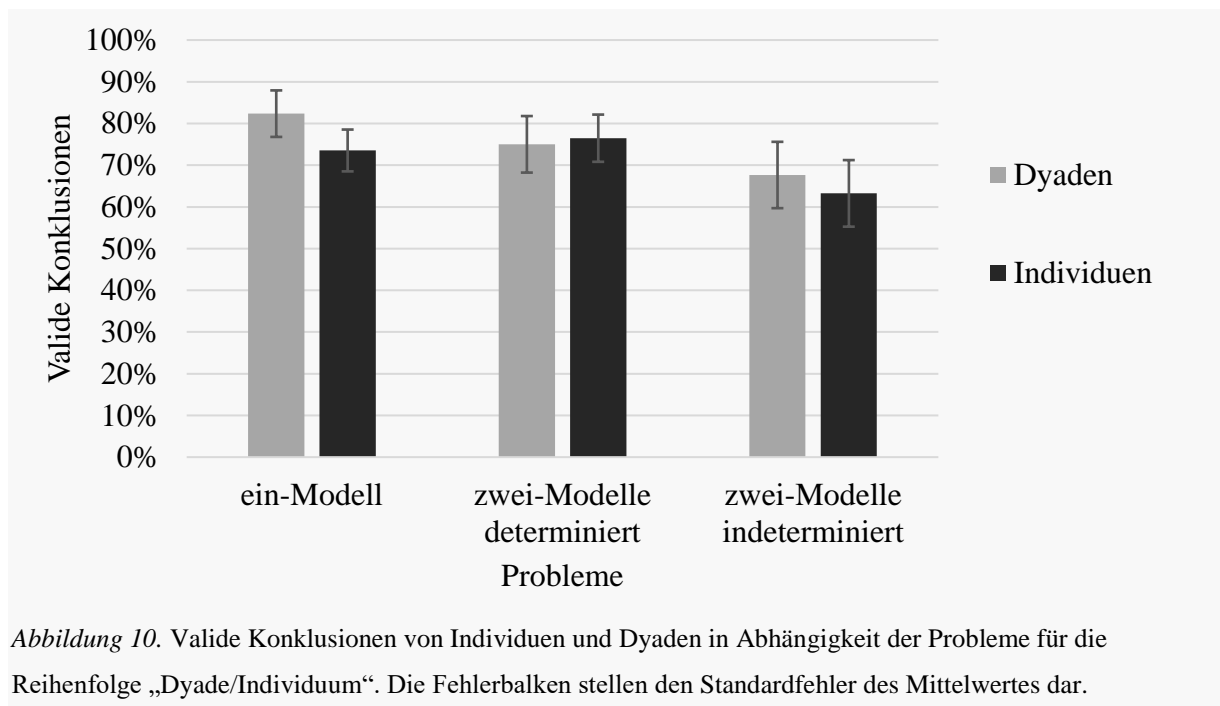


Abbildung 10. Valide Konklusionen von Individuen und Dyaden in Abhängigkeit der Probleme für die Reihenfolge „Dyade/Individuum“. Die Fehlerbalken stellen den Standardfehler des Mittelwertes dar.

Um die Unterschiede inferenzstatistisch zu testen, habe ich eine $2 \times 2 \times 3$ Varianzanalyse mit Messwiederholung mit dem Zwischensubjektfaktor „Reihenfolge“ (Individuum/Dyade \times Dyade/Individuum) und den Innersubjektfaktoren „Gruppe“ (Individuen \times Dyaden) und „Problem“ (ein-Modell \times zwei-Modelle determiniert \times zwei-Modelle indetermiert) durchgeführt. Die Analyse ergab einen signifikanten Haupteffekt für den Faktor „Gruppe“ ($F(1, 33) = 10.33, p = .003, \eta_p^2 = .238$) sowie eine marginal signifikante Interaktion „Gruppe \times Reihenfolge“ ($F(1, 33) = 3.53, p = .069, \eta_p^2 = .097$). Da der Haupteffekt vermutlich auf die Interaktion zurückzuführen ist, habe ich diese Interaktion weiter analysiert. Post-hoc t -Tests für verbundene Stichproben (Bonferroni korrigiertes Alpha-Niveau: $\alpha = .025$)

zeigen, dass in der Bedingung „Individuum/Dyade“ die Dyaden mit 83.80 % ($SD = 24.84$ %) insgesamt mehr valide Konklusionen ziehen als Individuen ($M = 62.73$ %, $SD = 22.71$ %; $t(53) = 4.80$, $p < .001$, $d = .89$). Bei der anderen Reihenfolge (Dyade/Individuum) zeigte sich kein Unterschied (Dyaden: $M = 75.00$ %, $SD = 28.28$ %; Individuen $M = 71.08$ %, $SD = 26.28$ %; $t(50) = 1.19$, $p = .238$, $d = .14$). Alle anderen Haupteffekte sowie Interaktionen wurden mit $p > .10$ nicht signifikant.

4.3.2.2 Entscheidungsschemata

Da ich in diesem Experiment dieselben Vpn zunächst alleine und dann in der Gruppe getestet habe, konnte ich nicht nur die drei theoretischen Modelle (*truth wins*, *equiprobability* und *error wins*) testen. Vielmehr erlaubte mir die Bedingung Individuum/Dyade die exakte Verteilung von individuellen Antworten mit den Gruppenantworten der gleichen Personen bei den $n = 72$ indeterminierten Problemen zu identifizieren. Dieses aus meinen empirischen Daten abgeleitete Modell nenne ich „*collaboration wins*“ Modell. Die Verteilung nach diesem Modell ist in Tabelle 12 dargestellt. Deskriptiv zeigt sich, dass selbst wenn nur ein Individuum zuvor die valide Konklusion ziehen konnte, die Gruppe in 82 % der Fälle gemeinsam eine valide Konklusion zog. Außerdem konnte die Dyade sogar noch sehr häufig (73 %) eine valide Konklusion ziehen, wenn dies keines der Individuen zuvor alleine gelang.

Tabelle 12

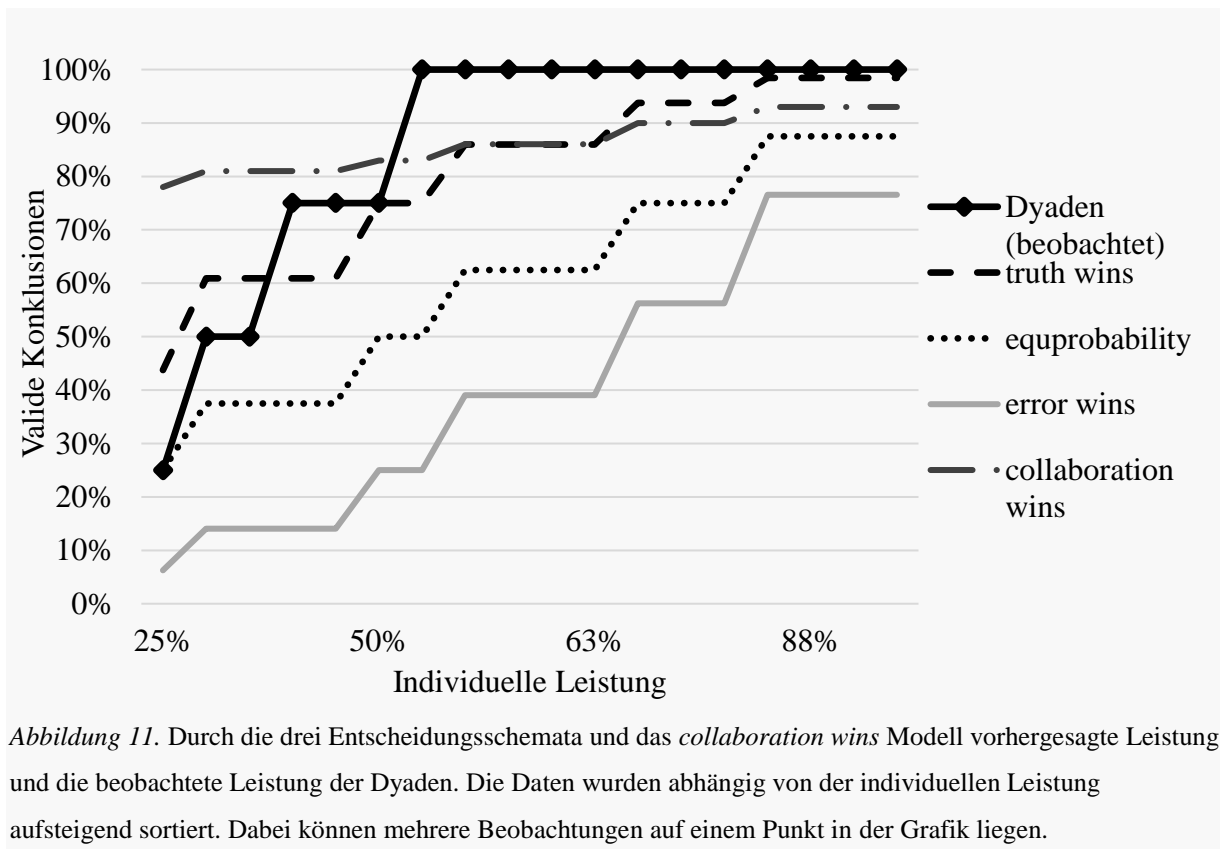
Collaboration wins Modell bei indeterminierten Problemen erstellt aus den Daten der Reihenfolge „Individuum/Dyade“

		Gruppenleistung	
r_s, r_n	n	C	I
2,0	27	96 %	4 %
1,1	34	82 %	18 %
0,2	11	73 %	27 %

Anmerkung: r_s = Anzahl der Mitglieder, welche die korrekte Lösung kannten; r_n = Anzahl der Mitglieder, welche die korrekte Lösung nicht kannten; n = Anzahl an Durchgängen; C = Wahrscheinlichkeit, dass die Gruppe die korrekte Lösung nennt; I = Wahrscheinlichkeit, dass die Gruppe die falsche Lösung nennt.

Die von dem *collaboration wins* vorhergesagte Leistung der Dyade ist zusammen mit den anderen drei Modellen (*truth wins*, *equiprobability*, *error wins*) in Abbildung 11 abgetragen. Man erkennt, dass die Leistung von Individuen (x-Achse) von $MIN = 25$ % bis

$MAX = 88\%$ validen Konklusionen reicht. Die Leistung der Dyaden reicht von $MIN = 25\%$ bis $MAX = 100\%$ validen Konklusionen. Deskriptiv zeigt sich, dass in vielen Fällen die Leistung der Dyaden die von allen Modellen vorhergesagte Leistung übertrifft. Der Kolmogorov-Smirnov Test zeigt, dass keines der Modelle die Leistung der Dyade eindeutig vorhersagen konnte (*collaboration wins*: $D_{max} = .667$, $Z(36) = 2.00$, $p = .001$; *truth wins*: $D_{max} = .667$, $Z(36) = 2.00$, $p = .001$; *equiprobability*: $D_{max} = .667$, $Z(36) = 2.00$, $p = .001$; *error wins*: $D_{max} = .667$, $Z(36) = 2.00$, $p = .001$). Jedes Modell hatte dabei exakt die gleichen Werte extremster Differenzen ($D_{max} = .667$). Deshalb stellte sich keines der Modelle zumindest als annähernd geeignetes Modell heraus. Um weiterhin zu überprüfen, ob das *collaboration wins* Modell einem der anderen drei Modelle ähnelt, habe ich dieses Modell mit den anderen drei Modellen verglichen. Der Kolmogorov-Smirnov Test zeigt, dass das *collaboration wins* Modell am meisten dem *truth wins* Modell ähnelt ($D_{max} = .389$, $Z(36) = 1.67$, $p = .131$). Deskriptiv zeigt sich dabei, dass das *collaboration wins* Modell vor allem für schlechtere individuelle Leistungen eine bessere Gruppenleistung bei indeterminierten Problemen vorhersagt. Die anderen beiden Modelle unterschieden sich eindeutig vom *collaboration wins* Modell ($p < .001$).



4.3.3 Diskussion

Im vorliegenden Experiment zeigte sich, dass Dyaden insgesamt mehr valide Konklusionen ziehen als Individuen. Dyaden können – unabhängig von der Schwierigkeit der Probleme – besser räumliche Schlussfolgerungen ziehen als Individuen. Deshalb stützen meine Daten die erste Hypothese, dass *Dyaden mehr valide Konklusionen ziehen als Individuen*. Wie in den vorherigen Experimenten sind die Dyaden in der Lage, die Prämissen besser zu verarbeiten und mehr valide Konklusionen zu ziehen. Die schwache Interaktion deutet jedoch darauf hin, dass dieser Unterschied nicht völlig unabhängig von der Reihenfolge war. Eine mögliche Erklärung für diesen Befund ist, dass die Vpn – unabhängig davon, ob sie alleine oder in der Gruppe gearbeitet haben – lediglich im ersten Durchgang etwas über die Aufgaben gelernt haben und deshalb im zweiten Durchgang besser wurden. Diese Annahme ist aber unwahrscheinlich. Wenn die Vpn lediglich dadurch besser wurden, dass sie die Probleme zweimal lösten, dann hätten die Individuen in der Bedingung „Dyade/Individuum“ mehr valide Konklusionen ziehen müssen als die Dyaden. Die deskriptiven Daten zeigen aber, dass die Individuen in dieser Bedingung die Leistung der Dyaden nicht übertreffen konnten. Deshalb weisen meine Ergebnisse darauf hin, dass die Personen in der Gruppe durch die Interaktion etwas lernten und dies für das spätere individuelle Denken nutzen konnten. Dass der Lerneffekt nicht einfach nur auf das doppelte Bearbeiten der Probleme zurückzuführen ist, kann auch durch mein Aufgabenmaterial gestützt werden. Um mögliche Lerneffekte zu minimieren, habe ich das Experiment so konzipiert, dass die Vpn in der zweiten Phase jeweils andere Terme (Obst vs. Werkzeug) präsentiert bekamen. Außerdem waren die Probleme in zufälliger Reihenfolge präsentiert. Es ist also sehr unwahrscheinlich, dass die Vpn die spezifischen Aufgaben lernten und sich im zweiten Durchgang an ihre logischen Schemata erinnerten. Wahrscheinlicher ist, dass die Vpn generell etwas über die Struktur der Probleme lernten und diese auf die nächsten Phasen übertragen konnten. Dieser generelle Transfer zeigte sich bereits bei dem Experiment von Maciejovsky und Budescu (2007) zum Denken mit der *Wason Selection Task*. Das Ziel meines Experimentes war jedoch nicht, explizit Lerneffekte zu testen. Um herauszufinden, in welcher Art das Lernen stattfindet, ist eine weitere Kontrollbedingung notwendig. In dieser Bedingung müssten die Vpn zweimal nacheinander alle 24 Probleme alleine lösen. Dann wäre es möglich, die Leistung des zweiten Durchgangs mit den individuellen Leistungen der Bedingung „Dyade/Individuum“ zu vergleichen. Hierdurch kann überprüft werden, ob die Vpn in der Gruppe gemeinsam Strategien entwickeln, die auf das individuelle Lösen der Probleme übertragen werden können. Meine Ergebnisse zeigen außerdem, dass die unterschiedlich schwierigen Probleme

(ein-Modell, zwei-Modelle determiniert und zwei-Modelle indeterminierte Probleme) keinen deutlichen Einfluss auf die Leistung von Individuen und Dyaden haben. Deshalb können meine Daten die zweite Hypothese, dass *Dyaden besonders bei indeterminierten Problemen mehr valide Konklusionen ziehen als Individuen* nicht stützen. Auch die Analyse der Entscheidungsschemata konnte in diesem Experiment keine eindeutigen Aufschlüsse über das Denken von Dyaden liefern. Kein Modell konnte die Leistung der Dyade adäquat vorhersagen. Daher kann die dritte Hypothese, dass *die Leistung der Dyade bei indeterminierten Problemen am besten durch das truth wins Schema vorhergesagt wird*, durch dieses Experiment nicht gestützt werden. Die Ergebnisse der Entscheidungsschemata liefern aber trotzdem weitere interessante Einsichten. Das *collaboration wins* Modell, welches ich aus der im Experiment beobachteten Leistung entwickelte habe, entsprach überwiegend dem *truth wins* Modell. Das *collaboration wins* Modell unterscheidet sich jedoch vor allem dann von den Vorhersagen des *truth wins* Modells, wenn die individuelle Leistung zuvor sehr gering war. Obwohl Individuen zuvor die Probleme nicht lösen konnten, konnten diese Personen in der Dyade in 72 % der Fälle trotzdem eine valide Konklusion ziehen. Dies ist ein Hinweis darauf, dass die überlegene Leistung der Dyade nicht ausschließlich auf das gegenseitige Überzeugen der korrekten Repräsentation zurückzuführen ist. Vielmehr scheint die Gruppe weitere kognitive Prozesse, die für das Lösen der Probleme hilfreich sind, zu fördern. Welche Prozesse dies im Detail sind, kann mein Experiment jedoch nicht eindeutig aufklären. Außerdem konnte ich mit meinen Daten einen Lerneffekt nicht eindeutig ausschließen, weshalb es sehr wichtig ist, die Vorhersagen des *collaboration wins* Modells in weiteren Experimenten zu testen. Da in meinen anderen zwei Experimenten zum Unbestimmtheitseffekt die Vpn bei den Beispieldurchgängen deutlich auf die korrekten Lösungen instruiert wurden, sind die Daten dieser Experimente nicht geeignet, um das Modell zu testen. In Zukunft müssen Experimente mit identischen Instruktionen und verschiedenen Reihenfolgen durchgeführt werden. Wenn das *collaboration wins* Modell ausschließlich durch einen Lerneffekt erklärt werden kann, dann sollte es ausschließlich die Gruppenleistung von Personen, die bereits zuvor individuell die Probleme lösten, vorhersagen können.

Zusammenfassend lässt sich sagen, dass in diesem Experiment die Leistung der Dyaden die Leistung von Individuen übertrifft. Dabei war der Unterschied jedoch nicht bei den besonders schwierigen indeterminierten Problemen am größten, sondern zeigte sich bei allen Problemen gleichermaßen. Das Experiment ermöglichte mir außerdem, das *collaboration wins* Modell zum kollaborativen räumlichen Schlussfolgern abzuleiten. Dieses Modell sagt vorher, dass sogar dann, wenn die Individuen nicht in der Lage sind, valide

Konklusionen zu ziehen, diese Personen in der Gruppe gemeinsam zu einer korrekten Lösung kommen können. Dieses Modell basiert aber ausschließlich auf den Beobachtungen dieses einen Experimentes. Deshalb ist es wichtig, das Modell zusammen mit den bereits bekannten Schemata aus der Literatur in weiteren Experimenten zu testen.

4.4 Zusammenfassung Kapitel 4

In diesem Kapitel habe ich die zweite Phase des Denkens, die Variation mentaler Modelle, getestet. Dazu habe ich den an Individuen ausführlich untersuchten Unbestimmtheitseffekt verwendet und geprüft, ob die Dyade die Annotation über mehr als eine Interpretation der Prämissen häufiger berücksichtigt als Individuen. Insgesamt sprechen meine Befunde für meine erste Hypothese, dass *Dyaden mehr valide Konklusionen ziehen als Individuen*. Im Mittel erreichten Dyaden über alle drei Experimente hinweg 13 % mehr valide Konklusionen als Individuen. Dieser Unterschied war deskriptiv mit 16 % bei den indeterminierten Problemen am größten (ein-Modell Probleme: 10 %; zwei-Modelle determinierte Probleme: 12 %). Trotz dieses deskriptiven Unterschiedes stellte sich in keinem der drei Experimente eindeutig heraus, dass Dyaden bei den besonders schwierigen indeterminierten Problemen am meisten profitieren. Deshalb sprechen die Befunde gegen meine zweite Hypothese, dass *Dyaden besonders bei indeterminierten Problemen mehr valide Konklusionen ziehen als Individuen*. Ein kritisches Problem bei allen drei Experimenten zum Unbestimmtheitseffekt ist, dass die für die Analyse verwendeten Beobachtungen nur sehr gering sind. So wurden in Experiment 5 zum Beispiel 70 Vpn getestet. Trotzdem gingen für die zwei Reihenfolgen jeweils nur 17 (Individuum/Dyade) bzw. 18 (Dyade/Individuum) Datensätze der Dyaden in die Auswertung ein. Der Grund dafür ist relativ trivial: Eine Dyade erzeugt (trotz zwei Personen) nur eine Antwort für jede Aufgabe. Dieses Problem wurde bereits von Davis (1996) als ein grundlegendes Hindernis für die Gruppenforschung bezeichnet. Davis betont dabei, dass die Stichprobengröße – obwohl zunächst nach einem banalen Problem klingend – ein nicht zu unterschätzendes Problem bei der Erforschung von Gruppen mit sich bringt. Zum einen steigt die notwendige Anzahl der notwendigen Vpn selbst bei einfachen experimentellen Designs um den Faktor der Gruppengröße und kann so schnell mehrere hundert Probanden erforderlich machen. Das zweite Problem ist, dass selbst bei einer Erscheinungsrate der einzelnen Vpn von 80 % und der Testung von kleinen Gruppen wie Dyaden, eine Gruppentestung in nur 64 % zu Stande kommt. Davis ging dabei so weit, dass durch diese Schwierigkeiten die Erforschung der Gruppen in der Sozialpsychologie vernachlässigt wurde.

Kießner und Ragni (2017) zeigten in ihrer umfassenden Meta-Analyse zum räumlichen Schlussfolgern, dass Individuen bei indeterminierten Problemen über alle Experimente hinweg im Mittel 52.0 % ($SD = 8.0$ %) valide Konklusionen zogen. In meinen Experimenten lagen Individuen mit 62.5 % ($SD = 29.1$ %) etwas über dieser Leistung. Dyaden hingegen erreichten in meinen Experimenten insgesamt 16 % mehr valide Konklusionen ($M = 78.4$ %, $SD = 25.3$ %) als Individuen. Ich nehme an, dass bei einer größeren Anzahl von Beobachtungen dieser deskriptive Unterschied mit großer Wahrscheinlichkeit durch die inferenzstatistische Analyse gestützt werden könnte.

Meine Experimente weisen auch unabhängig von der Anzahl der Beobachtungen eine Auffälligkeit auf. In Experiment 4 und Experiment 5 konnte ich den Unbestimmtheitseffekt weder bei Individuen noch bei Dyaden eindeutig zeigen. Besonders meine Befunde zum individuellen Denken decken sich deshalb nicht mit bisherigen Befunden zur Variation mentaler Modelle (Byrne & Johnson Laird, 1989; Kießner & Ragni, 2017; Ragni & Knauff, 2013). Insgesamt zeigen alle drei Experimente, dass Dyaden mehr valide Konklusionen ziehen als Individuen. Aber auch die Leistungen von Individuen waren in den Experimenten sehr gut. Es ist denkbar, dass die Aufgaben insgesamt nicht schwierig genug waren, um den Vorteil der Dyade deutlicher zeigen zu können. Ich nehme an, dass dies der Methodik meiner Experimente geschuldet ist. In den früheren Studien wurden die Prämissen überwiegend sequentiell, also nacheinander präsentiert. Ich habe mich jedoch für eine parallele Präsentation entschieden. Dies sollte besonders in Experiment 5 den Dyaden die Möglichkeit erleichtern, gemeinsam die Probleme zu diskutieren. In Experiment 3 und 4 wurden die Prämissen vor der Abfrage der Konklusion ausgeblendet. In diesen Experimenten zeigte sich noch ein schwacher Unbestimmtheitseffekt. In Experiment 5 – mit kompletter paralleler Präsentation – konnte ich den Unbestimmtheitseffekt nicht nachweisen. Dass die Art der Präsentation beim Unbestimmtheitseffekt nicht unbedeutend ist, konnte bereits Roberts (2000) zeigen. Auch in seinen Experimenten war der Unbestimmtheitseffekt am stärksten, wenn die Prämissen seriell präsentiert wurden. Roberts führt dies auf die Entlastung des Arbeitsgedächtnisses zurück. Dadurch, dass bei paralleler Präsentation alle Prämissen gleichzeitig präsentiert werden, wird das Arbeitsgedächtnis entlastet und den Personen fällt es leichter, mehrere Modelle zu berücksichtigen. Somit ist es bei zukünftigen Experimenten zur kollaborativen Variation mentaler Modelle wichtig, verschiedene Präsentationsarten als unabhängige Variable zu testen, um einen Mittelweg zwischen optimalen Gruppenbedingungen und kontrollierten Effekten zu finden.

In diesem Kapitel habe ich außerdem untersucht, durch welches soziale Kombinationsschema das Denken der Dyaden bei indeterminierten Problemen am besten beschrieben werden kann. In zwei von drei Experimenten konnte die Leistung der Dyade am besten durch das *truth wins* Entscheidungsschema vorhergesagt werden. Experiment 5 konnte diesen Befund nicht eindeutig stützen. Das aus den beobachteten Leistungen erstellte *collaboration wins* Modell stimmt jedoch stark mit dem *truth wins* Modell überein. Deshalb sprechen meine Ergebnisse insgesamt für die dritte Hypothese, dass *die Leistung der Dyade bei indeterminierten Problemen am besten durch das truth wins Schema vorhergesagt wird*. Die Personen in den Dyaden waren also in der Lage, sich gegenseitig von den korrekten Repräsentationen zu überzeugen. Dyaden konnten deskriptiv sogar häufig die vom *truth wins* Modell vorhergesagte Leistung übertreffen. Auch wenn beide Individuen bei indeterminierten Problemen keine valide Konklusion finden konnten, waren die Personen in der Lage, gemeinsam in der Dyade eine richtige Antwort zu finden. Deshalb unterscheidet sich das von mir gefundene *collaboration wins* Modell vor allem bei schlechter individueller Leistung vom *truth wins* Modell. Das *truth wins* Modell sieht anhand der individuellen Leistung nur eine geringe Chance, dass die Gruppe gemeinsam eine valide Konklusion zieht. Mein beobachtetes Modell nimmt jedoch eine höhere Gruppenleistung sogar dann an, wenn keines der Individuen eine valide Konklusion ziehen konnte. Das bedeutet, dass das *collaboration wins* Modell ebenfalls die von Moshman und Geil (1998) und Maciejovsky und Budescu (2007) deskriptiven Ergebnisse abbilden kann. Gruppen können auch dann korrekt schlussfolgern, wenn keines ihrer Individuen eine richtige Lösung finden konnte. Bei den von mir untersuchten Problemen zum räumlichen Schlussfolgern bedeutet dies, dass Personen in Dyaden auch dann die Annotation über weitere Interpretationen der Prämissen berücksichtigen, wenn die Individuen dies zuvor nicht getan haben.

Zusammenfassend konnten meine Experimente zeigen, dass die Gruppe den Individuen beim Denken überlegen ist. Obwohl Dyaden bei indeterminierten Problemen nicht eindeutig mehr valide Konklusionen ziehen konnten als Individuen, zeigten die deskriptiven Daten sowie die sozialen Entscheidungsschemata, dass die Dyade das Potential dazu haben kann. Die Personen in der Dyade konnten sich gegenseitig von den richtigen mentalen Modellen der Probleme überzeugen und bei den schwierigen indeterminierten Problemen sogar die vorhergesagte Leistung des *truth wins* Modells in vielen Fällen übertreffen. Wenn die Gruppenmitglieder sich jedoch gegenseitig von verschiedenen Repräsentationen überzeugen, führt dies unweigerlich dazu, dass mindestens eine Person in der Dyade mit inkonsistenten Informationen konfrontiert wird. Dann muss diese Person ihr ursprüngliches

Modell revidieren. Deshalb ist es sehr wichtig, die Revisionsprozesse räumlicher Überzeugungen auch im sozialen Kontext zu untersuchen. Diese soziale räumliche Überzeugungsrevision ist Gegenstand des nachfolgenden Kapitels.

Kapitel 5: Experimenteller Teil 3 - Soziale Revision mentaler Modelle

In diesem Kapitel habe ich in zwei Experimenten den sozialen Einfluss auf die Revision mentaler Modelle getestet. Beim individuellen Denken kann die Revision räumlicher Überzeugungen nicht durch Prinzipien wie der epistemischen Verwurzelung (Harman, 1986; Rott, 2001) oder dem *mismatch principle* (Johnson-Laird, et al., 2004) erklärt werden. Vielmehr stellte sich heraus, dass die räumliche Überzeugungsrevision durch die funktionelle Asymmetrie der Terme bestimmt wird (Bucher et al., 2013; Bucher & Nejasmic, 2012; Knauff et al., 2013; Krumnack et al., 2011). Dieses sogenannte LO-Prinzip stellte sich als sehr stabil heraus und konnte bisher kaum beeinflusst werden. Da in Gruppen beim kollaborativen Denken die Personen sich fortwährend von unterschiedlichen Repräsentationen überzeugen müssen, habe ich untersucht, ob das LO-Prinzip durch soziale Faktoren wie der Konformität oder das Quellenprinzip beeinflusst wird. Ich habe angenommen, dass das LO-Prinzip seltener gezeigt wird, wenn eine Majorität gegen diese Revision spricht. Den gleichen Einfluss habe ich erwartet, wenn die LO-Revision dazu führt, dass man die Aussage einer hoch vertrauenswürdigen Person verwerfen muss. Bei beiden Experimenten habe ich jedoch keine realen Gruppen getestet. Um den sozialen Einfluss gezielt manipulieren zu können, habe ich ein künstliches soziales Setting ohne direkte Interaktion geschaffen. Um jedoch das hinreichende Kriterium der Interaktion nach Shaw (1971) für Gruppen zu erfüllen, werden den Personen Informationen von anderen (fiktiven) Gruppenmitgliedern präsentiert. Maciejovsky und Budescu (2007) konnten beim Denken mit der *Wason Selection Task* zeigen, dass die Vpn auch ohne direkte Interaktion ein geteiltes mentales Modell der Probleme entwickelten und somit die Informationen der anderen Teilnehmer für ihre Konklusionen berücksichtigten. Auch in den Experimenten von Wolf et al. (2012) reichte es aus, den Personen eine Unterhaltung schriftlich zu präsentieren und die Vpn ließen sich von der Vertrauenswürdigkeit der Quelle bei ihren Revisionsentscheidungen beeinflussen. Deshalb nehme ich an, dass auch in meinen Experimenten die Präsentation von Informationen anderer – nicht real existierender – Gruppenmitglieder eine hinreichende minimale Interaktion darstellt und habe das experimentelle Paradigma von Wolf et al. (2012) für die folgenden Experimente übernommen.

Zusammenfassend die beiden Hypothesen dieses Kapitels:

Hypothese 1: Die Konformität hat einen Einfluss auf die Revisionsstrategie beim Revidieren räumlicher mentaler Modelle.

Hypothese 2: Das Quellenprinzip beeinflusst die Revisionsstrategie beim Revidieren räumlicher mentaler Modelle.

Die erste Hypothese habe ich in Experiment 6 und die zweite Hypothese mit Experiment 7 getestet. Alle Experimente dieses Kapitels wurden mit OpenSesame (Version 3.1, Mathôt et al., 2012) programmiert und durchgeführt. Die zufällige Zuordnung der Probanden habe ich mit randomizer.org (Version 4, Urbaniak & Plous, 2013) realisiert. Die Aufbereitung der Daten habe ich mit Microsoft Excel (Version 2016) durchgeführt und anschließend mit der Statistiksoftware SPSS (Version 24, IBM Corp., 2016) inferenzstatistisch ausgewertet. Für die Auswertung der relativen Häufigkeiten habe ich die angular transformierten Werte verwendet.

Die vorliegenden Experimente sind im Einklang mit den Standards der aktuellsten Version der Deklaration von Helsinki (World Medical Association, 2013) und wurden von einer lokalen Ethikkommission geprüft. Eine Ausnahme ist dabei, dass für die vorliegenden Experimente keine Pre-Registrierung nach §35 der Deklaration durchgeführt wurde.

5.1 Experiment 6: Konformität und räumliche Überzeugungsrevision

In diesem Experiment habe ich getestet, inwieweit die Konformität die Überzeugungsrevision beeinflusst. Wenn in einer Gruppe gegensätzliche Meinungen entstehen, führt dies zu starker kognitiver Dissonanz (Matz & Wood, 2005) und Personen sind bereit, ihre Meinungen zu Gunsten der Gruppenmeinung aufzugeben (Asch, 1951, 1956; Neubaum et al., 2018). Deshalb nehme ich an, dass dieser Konformitätsdruck auch auf die Revision räumlicher Überzeugungen wirkt. In diesem Experiment konnten die Vpn sich zwischen zwei logisch gleichwertigen Revisionsalternativen (LO- vs. RO-Revision) entscheiden und ich habe getestet, ob ein Konformitätsdruck auf der einen Seite dazu führen kann, dass das LO-Prinzip seltener gewählt wird, wenn die anderen Personen dieses Prinzip ablehnen. Auf der anderen Seite gehe ich davon aus, dass das LO-Prinzip häufiger verwendet wird, wenn sich die Majorität der Gruppe für dieses Prinzip entscheidet.

5.1.1 Methoden

5.1.1.1 Stichprobe

Die Stichprobe wurde mit Hilfe des Rundmail-Systems der Justus-Liebig-Universität akquiriert. Dabei wurden keine expliziten Ausschlusskriterien vordefiniert. An der Untersuchung nahmen $N = 22$ Vpn (20 weiblich) teil. Das mittlere Alter lag bei $M = 23.5$ ($SD = 3.1$) Jahren und reichte von $MIN = 19$ bis $MAX = 30$ Jahre. Für den Arbeitsaufwand erhielten die Vpn sogenannte Versuchspersonenstunden, die sie im Rahmen ihres Studiums benötigen. Als zusätzlicher Anreiz wurden unter allen Teilnehmern zwei Amazon-Gutscheine im Wert von je 25 € verlost.

5.1.1.2 Material

Das Material bestand aus 16 relational räumlichen drei-Term Aufgaben. Diese Aufgaben wurden anhand von vier Figuren (Tabelle 13) erstellt. Jede Figur führt zu dem Modell „ABC“. Den drei Termen wurden Objekte der Kategorie „Obst“ (Apfel, Birne, Mango, Kiwi, Orange) oder „Werkzeuge“ (Hammer, Säge, Bohrer, Feile, Meißel) vorab zugeordnet.

Tabelle 13

Vier Figuren für das Aufgabenmaterial. Jede der beiden Prämissen führt zu dem Modell „ABC“

	Figur 1	Figur 2	Figur 3	Figur 4
1. Prämisse	A links B	A links B	B rechts A	B rechts A
2. Prämisse	B links C	C rechts B	C rechts B	B links C
Modell	ABC			

Anmerkung: Den Termen A, B, und C wurden zufällig die Objekte der Kategorie „Obst“ (Apfel, Birne, Mango, Kiwi, Orange) oder Werkzeuge (Hammer, Säge, Bohrer, Feile, Meißel) zugeordnet.

Unter den beiden Prämissen stand ein in roter Schrift geschriebener Fakt. Dieser Fakt entsprach dabei immer dem Gegenteil der aus den beiden Prämissen resultierenden Konklusion. Das bedeutet, dass der Fakt entweder „C links A“ oder „A rechts C“ beinhaltete. Unter den drei Aussagen (Prämisse 1, Prämisse 2 und Fakt) wurden zwei Antwortalternativen in Form von Modellen dargeboten. Die eine Alternative entsprach dem Modell, welches durch die Revision nach dem LO-Prinzip entsteht. Die andere entsprach der RO-Revision. Neben den Modellen wurde zusätzlich in einer Aussage präsentiert, wie viele der anderen Teilnehmer sich für eine der beiden Alternativen entschieden haben. Dabei waren jeweils vier

Aufgaben so konstruiert, dass entweder keine, eine, zwei oder drei Personen das Modell, welches der LO-Revision entsprach, wählten. Bei dem Modell, welches der RO-Revision entsprach, stand entsprechend die gegenteilige Anzahl an Personen. Abbildung 12 stellt eine Aufgabe als Beispiel dar.

Aussage 1: Der Hammer ist links von der Feile.

Aussage 2: Die Feile ist links von dem Bohrer

Fakt: Der Hammer ist rechts von dem Bohrer

Welche Anordnung trifft eher zu?

Anordnung 1: Bohrer Hammer Feile *1 von 3 Probanden haben diese Anordnung gewählt.*

Anordnung 2: Feile Bohrer Hammer *2 von 3 Probanden haben diese Anordnung gewählt.*

Abbildung 12. Beispielaufgabe für den Fall, bei dem zwei von den Gruppenmitgliedern für die LO-Revision (Anordnung 2) gestimmt haben. Der Fakt war im Experiment in roter Schriftfarbe dargestellt.

5.1.1.3 Ablauf

Am Anfang des Experiments erhielten die Vpn Informationen zum Datenschutz und zur Dauer des Experiments. Danach unterschrieben die Vpn eine Einverständniserklärung. Anschließend wurde die Instruktion auf dem Monitor präsentiert. Den Vpn wurde mitgeteilt, dass sie sich die Anordnung, die aus den zwei Prämissen entsteht, mental als zusammenhängende Anordnung vorstellen sollten. Außerdem wurde ihnen gesagt, dass ein zusätzlicher Fakt in roter Schrift präsentiert wird. Es wurde betont, dass dieser Fakt immer wahr ist und deshalb immer berücksichtigt werden muss. Deshalb sollten sich die Vpn entscheiden, welche der beiden (revidierten) Anordnungen für sie am ehesten zutrifft. Alle Informationen wurden parallel auf einem Bild präsentiert. Die Vpn wurden außerdem darüber informiert, dass sie gemeinsam mit drei weiteren Personen in einem gruppenähnlichen Setting an dem Experiment teilnehmen. Dabei wurde ihnen gesagt, dass die anderen Vpn in separaten Laborräumen sitzen und zur gleichen Zeit dieselben Aufgaben lösen würden. Die Vpn wurden instruiert, dass sie auf ihrem Monitor die Antworten der anderen Vpn präsentiert bekommen.

Sie erhielten dabei die Information, dass die anderen drei Vpn ihre Antworten immer als erstes geben würden. Somit bekamen die echten Vpn immer die Entscheidungen von allen drei anderen Teilnehmern präsentiert. Damit die Vpn im Glauben bestärkt wurden, dass auch andere Personen an dem Experiment teilnahmen, lief der Versuchsleiter, während die Vpn die demografischen Daten ausfüllten, zu anderen Räumen und tat so, als ob er mit den weiteren Vpn redete und Fragen beantwortete. Vor dem Hauptexperiment durchliefen alle Vpn vier Übungsdurchgänge, bei denen sie Fragen stellen konnten. Die Antworten wurden mit der Taste 1 (erste Anordnung) oder Taste 2 (zweite Anordnung) auf der Tastatur gegeben. Dabei war ausbalanciert, ob das erste oder das zweite Modell der LO-Revision entsprach. Nach jeder Aufgabe erschien ein Pausenfenster und die Vpn konnten selbstständig entscheiden, wann sie fortführen. Nachdem die Vpn die 16 Aufgaben des Hauptexperimentes beantwortet hatten, wurden sie über das Ziel des Experiments aufgeklärt. Das gesamte Experiment dauerte 30 Minuten.

5.1.1.4 Design

Das Experiment folgt einem einfaktoriellen Design mit dem vierstufigen Innersubjektfaktor „Konformität“ (keine Person \times eine Person \times zwei Personen \times drei Personen). Abhängige Variable ist die Wahl der LO-Revision. Das experimentelle Design ist als Überblick mit der Anzahl der Aufgaben in Tabelle 14 dargestellt.

Tabelle 14

Experimentelles Design mit $N = 22$ Versuchspersonen. Abhängige Variable ist die LO-Revision

Konformität (Innersubjektfaktor)			
keine Person	eine Person	zwei Personen	drei Personen
4 Aufgaben	4 Aufgaben	4 Aufgaben	4 Aufgaben

5.1.2 Ergebnisse

Aus den Antworten der vier Aufgaben zu jeder Faktorstufe habe ich eine relative Häufigkeit der Wahl zur LO-Revision gebildet. Deskriptiv zeigt sich, dass sich die Vpn in 35.23 % ($SD = 22.70$ %) der Fälle für die LO-Revision entschieden haben, wenn keine andere Person diese wählte. Wenn eine Person für die LO-Revision war, entschieden sich die Vpn zu 45.45 % (SD

= 23.95 %) ebenfalls für diese Revision. Bei zwei Personen war die LO-Entscheidung deskriptiv nur etwas höher 48.86 % ($SD = 22.46$ %). Wenn sich alle drei anderen Personen für LO-Revision entschieden, wählten die Vpn zu 68.18 % ($SD = 25.80$ %) diese Revision. Die deskriptiven Daten sind zur Veranschaulichung in Abbildung 13 abgetragen.

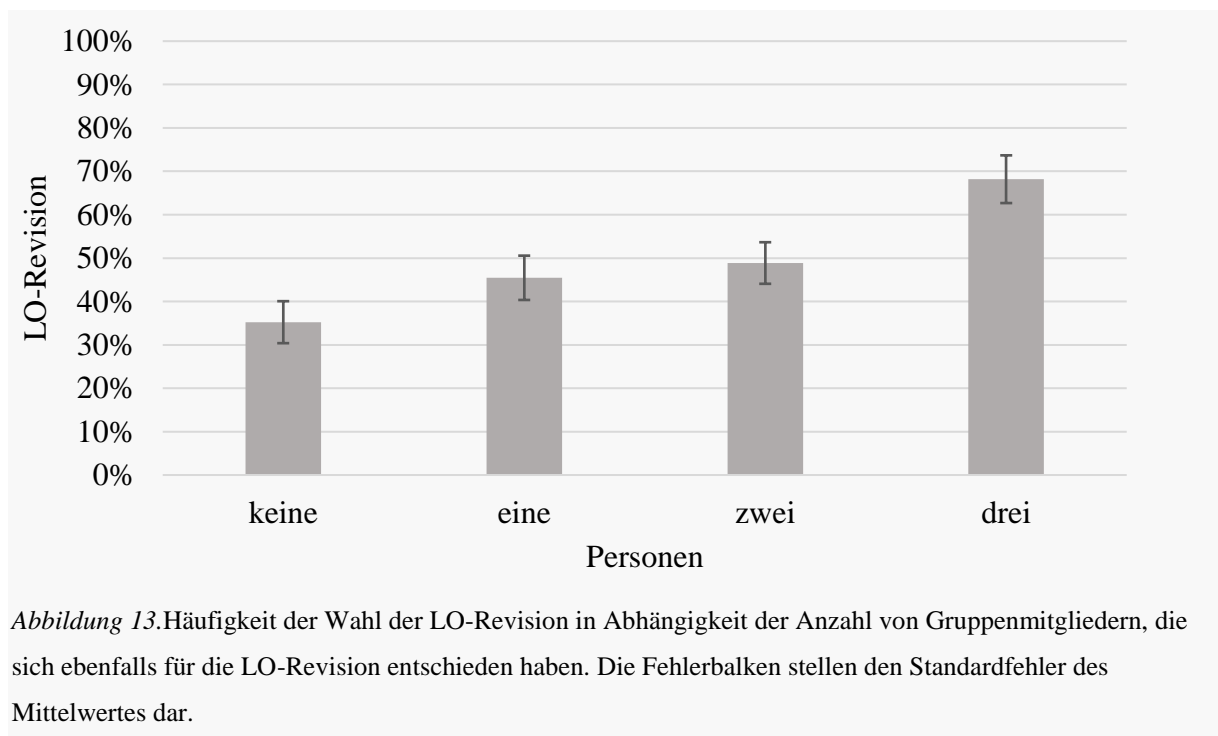


Abbildung 13. Häufigkeit der Wahl der LO-Revision in Abhängigkeit der Anzahl von Gruppenmitgliedern, die sich ebenfalls für die LO-Revision entschieden haben. Die Fehlerbalken stellen den Standardfehler des Mittelwertes dar.

Um die deskriptiven Unterschiede inferenzstatistisch zu prüfen, habe ich eine einfaktorielle Varianzanalyse mit Messwiederholung für den Innersubjektfaktor „Konformität“ (keine Person \times eine Person \times zwei Personen \times drei Personen) durchgeführt. Die Analyse ergab einen Haupteffekt für den Faktor „Konformität“ ($F(3, 63) = 6.82, p < .001, \eta_p^2 = .244$). Um diesen Haupteffekt genauer zu erklären, habe ich sechs post hoc t -Tests für verbundene Stichproben durchgeführt. Um dabei einer Alpha-Fehler Kumulation durch multiples Testen entgegenzuwirken, habe ich das Alpha-Niveau nach Bonferroni korrigiert (korrigiertes $\alpha = .008$). Wenn keine der Personen sich für das LO entschied, wählten die Vpn das LO-Prinzip signifikant seltener, als wenn alle Personen dieses Prinzip wählten ($t(21) = 3.40, p = .003, d = .1.36$). Zwischen einer Person und allen Personen zeigte sich ein marginal signifikanter Unterschied ($t(21) = 2.82, p = .010, d = .91$). Der Unterschied zwischen zwei Personen und allen Personen war ebenfalls signifikant ($t(21) = 3.26, p = .004, d = .80$). Alle anderen Vergleiche wurden mit $p > .10$ nicht signifikant.

5.1.3 Diskussion

Wenn keine Person einer fiktiven Gruppe beim Revidieren räumlicher Überzeugungen das Modell der LO-Revision wählte, dann entschieden sich die Vpn in nur 35 % der Fälle ebenfalls für diese Revision. Wenn sich jedoch alle Personen für die LO-Revision entschieden, dann verstärkte sich der LO-Effekt und die Vpn wählten zu 68 % diese Revision. Dieser Befund stützt meine Hypothese, dass *die Konformität einen Einfluss auf die Revisionsstrategie beim Revidieren räumlicher mentaler Modelle hat*. Personen lassen sich beim Revidieren ihrer mentalen Modelle durch die Antworten von anderen Personen beeinflussen. Dieser Befund ist aus zwei Perspektiven sehr wichtig für die Erforschung von Gruppen. Auf der einen Seite reihen sich meine Befunde in die schon sehr früh von Asch (1951, 1956) gefundenen und bis heute gestützten Ergebnisse ein (Neubaum et al., 2018). Obwohl die Revisionsstrategien logisch absolut gleichwertig sind, lassen sich Personen von der Majorität beeinflussen. Auf der anderen Seite zeigt das vorliegende Experiment, wie wichtig soziale Einflüsse auch für den experimentellen Bereich des kollaborativen räumlichen Schlussfolgerns sein können. Wenn man Experimente mit mehr als zwei Personen durchführen möchte, können schnell Majoritäten entstehen. Diese Majoritäten sind aber nur dann hilfreich, wenn sich die meisten Personen einer Gruppe für eine valide Konklusion entschieden haben. Wenn sich die Majorität für eine falsche Antwort entscheidet, kann dies schnell dazu führen, dass die Gruppe die Leistung von Individuen beim Denken nicht übertreffen kann. Wenn man also größere Gruppen untersuchen möchte, sollte man Möglichkeiten schaffen, dass jede Person gleichermaßen ihre Konklusion verteidigen und der Gruppe mitteilen kann.

Interessant bei den vorliegenden Ergebnissen ist auch, dass in den mittleren Kategorien (eine Person und zwei Personen) sich die Vpn in nur der Hälfte der Fälle für die LO-Revision entschieden haben. Diese Ergebnisse geben uns einen Hinweis darauf, dass nur extreme Gruppenmeinungen die räumliche Überzeugungsrevision eindeutig beeinflussen. Auffällig bei den Ergebnissen ist jedoch, dass die LO-Revision insgesamt nicht sehr häufig gewählt wurde. In bisherigen Studien zur räumlichen Überzeugungsrevision entschieden sich die Vpn in bis zu 85 % der Fälle für die LO-Revision (Bucher et al., 2011; Knauff et al., 2013). In meinem Experiment lag das Maximum bei 68 %. In den mittleren Kategorien wählten die Vpn eher zufällig die LO- bzw. die RO-Revision. Dieser Befund könnte ein Hinweis darauf sein, dass das *mismatch principle* nach Johnson-Laird et al. (2004) die Revisionsstrategie in diesem sozialen Setting besser vorhersagt als das LO-Prinzip. Das *mismatch principle* besagt, dass bei logisch gleichwertigen Revisionsalternativen die Vpn

zufällig eine Revision wählen. Diese Vorhersage stimmt mit den Antworten der Vpn in meinem Experiment nur dann überein, wenn sich nicht alle anderen Personen für beziehungsweise gegen eine bestimmte Antwortalternative entschieden. Es könnte also sein, dass die Vpn durch das Wissen, dass sie gemeinsam mit anderen Personen an den Aufgaben arbeitete sehr stark daran interessiert waren, die logisch beste Antwort zu finden. Dabei könnten sie erkannt haben, dass beide Antwortalternativen (LO- bzw. RO-Revision) logisch absolut gleichwertig sind. Somit antworteten sie in den mittleren Konformitätsausprägungen nicht nach dem LO-Prinzip, sondern – wie vom *mismatch principle* vorhergesagt – rein zufällig. Wenn sich jedoch eine Majorität für oder gegen eine bestimmte Revision bildete, überschrieb die Konformität diese logische Gleichwertigkeit und die Vpn übernahmen die Antwort der Majorität.

Insgesamt kann das vorliegende Experiment zeigen, dass der soziale Einfluss der Konformität auch beim Revidieren mit räumlichen mentalen Modellen ein wichtiger Einflussfaktor sein kann. Im nächsten Experiment habe ich untersucht, ob die Vertrauenswürdigkeit einer Quelle ebenfalls das LO-Prinzip beeinflusst.

5.2 Experiment 7: Quellenprinzip und räumliche Überzeugungsrevision

Wir wissen aus den Experimenten von Wolf et al. (2012) zur Überzeugungsrevision mit konditionalen Aufgaben, dass die Vertrauenswürdigkeit einer Quelle einen starken Einfluss darauf haben kann, welche Prämisse Personen verwerfen. Wenn eine Prämisse von einer niedrig vertrauenswürdigen Quelle geäußert wurde, dann waren die Probanden eher bereit, diese Prämisse aufzugeben. Bei der räumlichen Überzeugungsrevision verwerfen die Probanden immer die Prämisse, die nicht mit der LO-Revision übereinstimmt. In diesem Experiment habe ich getestet, ob das Quellenprinzip das LO-Prinzip beeinflussen kann. Ich habe geprüft, ob die Vpn das LO-Prinzip seltener verwenden, wenn diese Revisionsstrategie bedeuten würde, dass sie dafür die Aussage einer vertrauenswürdigen Quelle aufgeben müssen. Zudem habe ich getestet, ob das LO-Prinzip verstärkt wird, wenn durch diese Revision die Aussage einer wenig vertrauenswürdigen Quelle aufgegeben wird. Ich habe also angenommen, dass eine niedrig vertrauenswürdige Quelle das LO-Prinzip verstärkt und bei einer hoch vertrauenswürdigen Quelle das LO-Prinzip seltener gezeigt wird.

Um die Vertrauenswürdigkeit für mein Hauptexperiment manipulieren zu können, habe ich eine Vorstudie, in der verschiedene Berufe nach ihrer Vertrauenswürdigkeit beurteilt wurden, durchgeführt. Anhand der Ergebnisse der Vorstudie habe ich daraufhin verschiedene Aufgaben für das Hauptexperiment konstruiert.

5.2.1 Vorstudie

Die Vorstudie wurde mit Hilfe der Online Plattform SoSciSurvey (Version 3.1.06, Leiner, 2019) durchgeführt. Dabei wurden keine expliziten Ausschlusskriterien vordefiniert. An der Umfrage nahmen $N = 212$ (176 weiblich) Personen teil. Das mittlere Alter lag bei $M = 23.68$ ($SD = 4.73$) Jahren und reichte von $MIN = 18$ bis $MAX = 61$ Jahre. Der Fragebogen umfasste 33 deutsche Übersetzungen der Berufe aus der Studie von Wolf et al. (2012). Jeder Beruf sollte auf einer Likert-Skala von 1 (*überhaupt nicht vertrauenswürdig*) bis 7 (*sehr vertrauenswürdig*) bewertet werden. Die Reihenfolge der Berufe war für jede Vpn randomisiert dargeboten. Die 22 für das Hauptexperiment ausgewählten Berufe sind in Tabelle 15 abgetragen. Um die Berufe in die verschiedenen Vertrauenswürdigkeiten (niedrig, mittel, hoch) einzuordnen, habe ich den Median ($Md = 4.48$) gebildet und sieben Berufe aus dem obersten Quartil (hoch vertrauenswürdig), sieben aus dem untersten Quartil (niedrig vertrauenswürdig) und acht Berufe, die dem Median am nächsten waren (mittel vertrauenswürdig) ausgewählt. T -Tests für verbundene Stichproben ergaben, dass die hoch vertrauenswürdigen Berufe ($M = 5.58$, $SD = 0.25$) signifikant vertrauenswürdiger beurteilt wurden, als mittel ($M = 4.50$, $SD = 0.15$; $t(6) = 20.53$, $p < .001$, $d = 5.24$) und niedrig vertrauenswürdige Berufe ($M = 3.07$, $SD = 0.32$; $t(6) = 45.72$, $p < .001$, $d = 8.74$). Mittel vertrauenswürdige Berufe wurden signifikant vertrauenswürdiger beurteilt als niedrig vertrauenswürdige Berufe ($t(6) = 19.66$, $p < .001$, $d = 5.72$).

Tabelle 15

Für das Hauptexperiment 22 ausgewählte Berufe mit den mittleren Bewertungen der Vertrauenswürdigkeit

Hoch	Mittel	Niedrig
Feuerwehrmann ($M = 5.89$)	Metzger ($M = 4.68$)	Minister ($M = 3.51$)
Tierarzt ($M = 5.74$)	Fotograf ($M = 4.67$)	Bundestagsabgeordneter ($M = 3.42$)
Richter ($M = 5.72$)	Student ($M = 4.51$)	Spion ($M = 3.24$)
Pilot ($M = 5.72$)	Sachbearbeiter ($M = 4.49$)	Börsenmakler ($M = 2.95$)
Psychologe ($M = 5.40$)	Jäger ($M = 4.48$)	Versicherungsvertreter ($M = 2.86$)
Polizist ($M = 5.35$)	Friseur ($M = 4.39$)	Immobilienmakler ($M = 2.85$)
Entwicklungshelfer ($M = 5.22$)	Berater ($M = 4.26$)	Gebrauchtwagenhändler ($M = 2.66$)
	Radiomoderator ($M = 4.19$)	

5.2.2 Methoden

5.2.2.1 Stichprobe

Die Stichprobe wurde mit Hilfe des Rundmail-Systems der Justus-Liebig-Universität akquiriert. Dabei wurden keine expliziten Ausschlusskriterien vordefiniert. An der Untersuchung nahmen $N = 15$ Vpn (13 weiblich) teil. Das mittlere Alter lag bei $M = 21.9$ ($SD = 2.9$) Jahren und reichte von $MIN = 19$ bis $MAX = 29$ Jahre. Für den Arbeitsaufwand erhielten die Vpn sogenannte Versuchspersonenstunden, die sie im Rahmen ihres Studiums benötigen. Als zusätzlicher Anreiz wurden unter allen Teilnehmern zwei Amazon-Gutscheine im Wert von je 25€ verlost.

5.2.2.2 Material

Das Material bestand aus zwölf relational räumlichen drei-Term Aufgaben. Diese Aufgaben wurden anhand von den vier Figuren, die ich bereits in Experiment 6 zur Konformität verwendet habe, erstellt. Jede Figur führt zu dem Modell „ABC“. Den drei Termen wurden Objekte der Kategorie „Obst“ (Apfel, Birne, Mango, Kiwi, Orange) oder „Werkzeuge“ (Hammer, Säge, Bohrer, Feile, Meißel) vorab zugeordnet. Vor jeder der beiden Prämissen stand jeweils der Name eines Berufs. Diese Berufe waren so gepaart, dass entweder ein Beruf aus der Kategorie „hoch vertrauenswürdig“ und der andere Beruf aus der Kategorie „niedrig vertrauenswürdig“ stammte oder beide Berufe „mittel vertrauenswürdig“ waren. Die mittleren vertrauenswürdig Paare wurden verwendet, um das LO-Prinzip ohne einen direkten Einfluss der Vertrauenswürdigkeit zu testen (Bedingung „mittel“). Bei der hoch/niedrig Paarung war in 50 % der Fälle ein hoch vertrauenswürdiger Beruf der Prämisse zugeordnet (Bedingung „hoch“), die durch die LO-Revision verworfen wird. In der anderen Hälfte war die Quelle dieser Prämisse ein niedrig vertrauenswürdiger Beruf (Bedingung „niedrig“). Dabei war ausbalanciert, ob die erste oder die zweite Prämisse durch die LO-Revision verworfen wurde. Weiterhin enthielt das Material einen in roter Schriftfarbe geschriebenen Fakt. Dieser Fakt entsprach dabei immer dem Gegenteil der aus den beiden Prämissen resultierenden Konklusion. Das heißt, dass der Fakt entweder „C links A“ oder „A rechts C“ beinhaltete. Abbildung 14 stellt eine Beispielaufgabe für den Fall, dass durch das LO-Prinzip die Prämisse der vertrauenswürdigen Quelle (Tierarzt) verworfen wird, dar.

Unterhaltung zwischen einem **Tierarzt** und einem **Spion**.

Tierarzt: „Die Säge ist links von dem Bohrer.“

Spion: „Der Meißel ist rechts von dem Bohrer.“

Säge Bohrer Meißel oder **Meißel Bohrer Säge**

(korrekt)

(inkorrekt)

Fakt: Die Säge ist rechts von dem Meißel

Bohrer Meißel Säge oder **Meißel Säge Bohrer**

(LO-Revision)

(RO-Revision)

Abbildung 14. Beispielaufgabe bei der die Prämisse der vertrauenswürdigen Quelle (Tierarzt) verworfen wird. Jeder Satz wurde im Experiment sequentiell dargeboten.

5.2.2.3 Ablauf

Vor Beginn des Experiments wurden die Vpn über die Datenschutzrichtlinien informiert und haben eine Einverständniserklärung unterschrieben. Das gesamte Experiment fand an einem handelsüblichen Computer statt. Die Vpn konnten dabei selbstständig mit Druck auf die Leertaste entscheiden, wann sie fortfahren wollten. Eine Aufgabe folgte immer dem gleichen Schema. Die Vpn bekamen zuerst die Information darüber, dass sie ein Gespräch von zwei Personen aus unterschiedlichen Berufsgruppen präsentiert bekommen. Den Vpn wurde gesagt, dass jede Person jeweils eine Aussage über die Anordnung von zwei Objekten tätigt. Dabei wurde betont, dass der Inhalt der Aussagen nicht notwendigerweise wahr sein müsse, sie sich aber trotzdem eine zusammenhängende Anordnung der drei in den Aussagen genannten Objekten merken sollten. Nachdem die Vpn die zweite Prämisse gelesen hatten, wurden ihnen zwei mögliche Anordnungen präsentiert. Eine Anordnung entsprach dabei der aus den Prämissen resultierenden Anordnung (korrektes Modell). Die andere Anordnung war das Invers der korrekten Anordnung (inkorrektes Modell). Die Vpn wählten mit der linken bzw. rechten Pfeiltaste, welches der Modelle ihrer gemerkten Anordnung entsprach. Diese Abfrage sollte sicherstellen, dass die Vpn ein korrektes mentales Modell der Prämissen für die Revisionsphase konstruiert hatten. Dabei war die Position des korrekten Modells (links oder rechts) über das Experiment ausbalanciert. Nachdem die Vpn eine Antwort gegeben hatten, wurde der Fakt in roter Schrift präsentiert. Den Vpn wurde gesagt, dass dieser Fakt immer wahr sei und immer berücksichtigt werden müsse. Nachdem sie den Fakt gelesen hatten, begann die Revisionsphase. Den Vpn wurden auf der nächsten Seite zwei mögliche

Revisionsalternativen präsentiert. Eine der Alternativen entsprach dabei der Revision nach dem LO-Prinzip. Die andere Alternative entsprach der RO-Revision. Wieder konnten die Vpn mit der linken bzw. rechten Pfeiltaste entscheiden, welche Anordnung sie wählen wollten. Dabei wurden die Art der Revision (LO- bzw. RO-Alternative) und die Antwortzeiten aufgezeichnet. Die Position der LO-Alternative (links oder rechts) war wieder über das Experiment ausbalanciert.

Nachdem die Vpn eine Wahl getroffen hatten, wurde überprüft, ob sie die Quellen, die vor den Prämissen präsentiert waren, verarbeitet hatten. Dazu wurden die Vpn gefragt, zwischen welchen Berufen die Unterhaltung stattfand. Ihnen wurden zwei Berufspaare präsentiert. Ein Paar entsprach den beiden Personen aus der Unterhaltung und das andere Paar waren zufällig ausgewählte Berufe, die nicht in der Unterhaltung vorkamen. Dabei war wieder die Position des richtigen Berufspaares (links oder rechts) über das Experiment ausbalanciert. Nachdem die Vpn die Frage beantwortet hatten, konnten sie eine Pause machen und selbstständig entscheiden, wann sie fortfuhren. Vor dem Hauptexperiment bearbeiteten die Vpn zwei Probleme als Übungsdurchgang. Nachdem die Vpn alle zwölf Aufgaben des Hauptexperiments gelöst hatten, haben die Vpn den Fragebogen zu der Vertrauenswürdigkeit aus der Vorstudie zu den Berufen auf einer Likert-Skala von 1 (*überhaupt nicht vertrauenswürdig*) bis 7 (*sehr vertrauenswürdig*) beantwortet. Danach wurden sie über das Ziel des Experiments aufgeklärt und verabschiedet. Das gesamte Experiment dauerte 45 Minuten.

5.2.2.4 Design

Das Experiment folgte einem einfaktoriellen Design mit dem dreistufigen Innersubjektfaktor „Vertrauenswürdigkeit“ (niedrig \times mittel \times hoch). Abhängige Variablen waren die Wahl der LO-Revision und die Antwortzeit für die Revision. Das experimentelle Design ist als Überblick mit der Anzahl der Aufgaben in Tabelle 16 dargestellt.

Tabelle 16

Experimentelles Design mit $N = 15$ Versuchspersonen. Abhängige Variablen LO-Revision und die Antwortzeiten

Vertrauenswürdigkeit der Quelle (Innersubjektfaktor)		
niedrig	mittel	hoch
4 Aufgaben	4 Aufgaben	4 Aufgaben

5.2.3 Ergebnisse

Von allen $N = 180$ Durchgängen des Hauptexperimentes wurden die Durchgänge aus der Auswertung entfernt, bei denen die Vpn ein falsches Modell konstruiert oder die Frage zu den Berufen aus der Unterhaltung falsch beantwortet haben. Insgesamt mussten $n = 17$ Durchgänge aus der Auswertung entfernt werden. Somit wurden nur die Daten ausgewertet, bei denen sicher war, dass die Vpn (1) ein korrektes mentales Modell als Grundlage für ihre Revision verwendet und (2) die Berufe als Quellen der Prämissen korrekt verarbeitet haben. Aus den verbleibenden $n = 163$ Durchgängen wurden die relativen Häufigkeiten der LO-Revision sowie mittleren Antwortzeiten für die Analyse gebildet.

5.2.3.1 Manipulationscheck

Um zu prüfen, dass die Berufe im Hauptexperiment unterschiedlich vertrauenswürdig empfunden wurden, habe ich für die Daten des Fragebogens aus dem Hauptexperiment die gleiche Analyse wie bereits in der Vorstudie durchgeführt. Die beurteilten Vertrauenswürdigkeiten für alle Berufe aus der Vorstudie und dem Hauptexperiment sind Anhang C zu entnehmen. Der Median war im Hauptexperiment mit $Md = 4.53$ mit dem Median ($Md = 4.48$) zur Vorstudie vergleichbar. Bis auf zwei Berufe (Polizist: $M = 4.67$ $SD = 1.72$; Entwicklungshelfer: $M = 4.67$, $SD = 1.54$) konnten alle eindeutig den gleichen Kategorien (niedrig, mittel, hoch) zugeordnet werden. Der Polizist und der Entwicklungshelfer waren nicht mehr eindeutig in der Kategorie „hoch vertrauenswürdig“. Vielmehr lagen diese Berufe am oberen Ende der mittleren Kategorie. Insgesamt zeigte sich, dass wie in der Vorstudie hoch vertrauenswürdige Berufe ($M = 5.37$, $SD = 0.52$) signifikant vertrauenswürdiger beurteilt wurden als mittel ($M = 4.50$, $SD = 0.13$; $t(6) = 5.87$, $p = .001$, $d = 2.30$) und niedrig vertrauenswürdige Berufe ($M = 3.10$, $SD = 0.48$; $t(6) = 36.72$, $p < .001$, $d = 4.54$). Mittel vertrauenswürdige Berufe wurden wieder signifikant vertrauenswürdiger beurteilt als niedrig vertrauenswürdige Berufe ($t(6) = 10.56$, $p < .001$, $d = 3.98$).

5.2.3.2 Revisionen

Deskriptiv zeigt sich, dass sich die Vpn in 71.67 % ($SD = 32.40$ %) der Fälle für die LO-Revision entschieden haben, wenn die Quelle niedrig vertrauenswürdig war. Wenn die Quelle einer mittleren Vertrauenswürdigkeit entsprach, entschieden sich die Vpn zu 65.00 % ($SD = 34.39$ %) für die LO-Revision. Bei einer hoch vertrauenswürdigen Quelle verwendeten die Vpn in 70.56 % ($SD = 20.86$ %) der Fälle die LO-Revision. Die deskriptiven Daten sind zur Veranschaulichung in Abbildung 15 abgetragen.

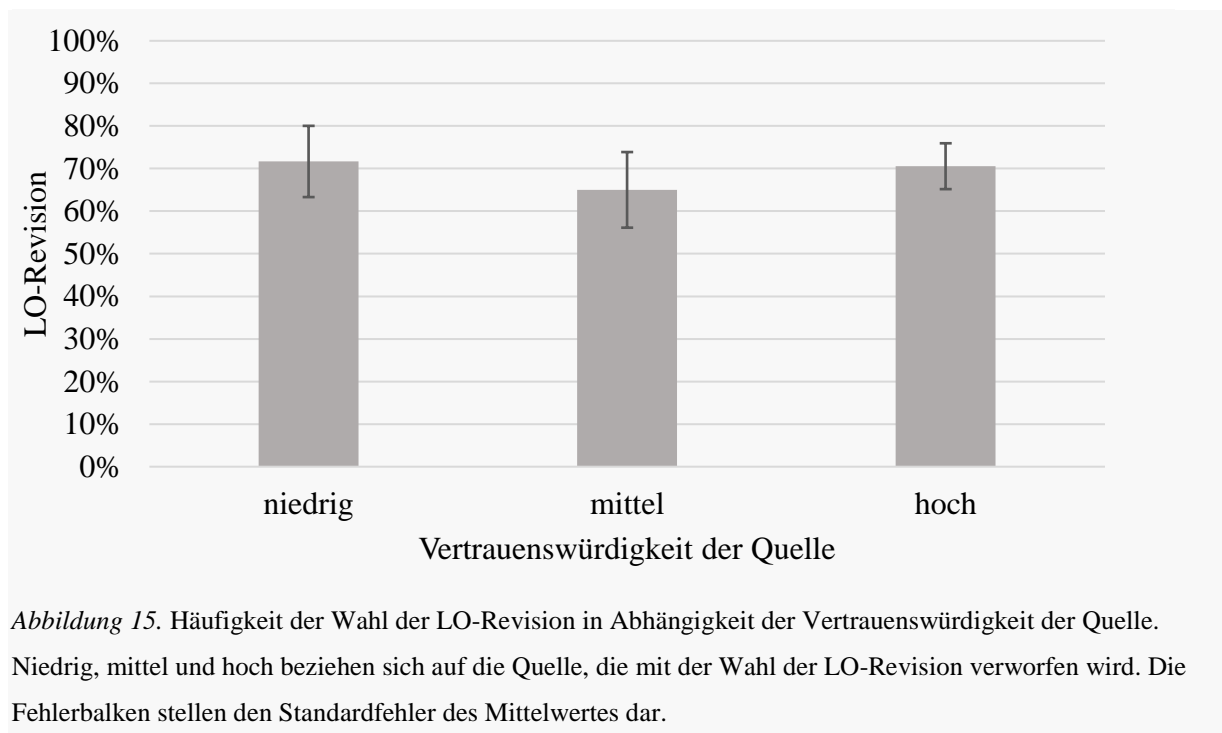


Abbildung 15. Häufigkeit der Wahl der LO-Revision in Abhängigkeit der Vertrauenswürdigkeit der Quelle. Niedrig, mittel und hoch beziehen sich auf die Quelle, die mit der Wahl der LO-Revision verworfen wird. Die Fehlerbalken stellen den Standardfehler des Mittelwertes dar.

Um die deskriptiven Unterschiede inferenzstatistisch zu prüfen, habe ich eine einfaktorielle Varianzanalyse mit Messwiederholung für den Innersubjektfaktor „Vertrauenswürdigkeit“ (niedrig \times mittel \times hoch) durchgeführt. Die Analyse ergab keinen Haupteffekt der LO-Revisionen für den Faktor „Vertrauenswürdigkeit“ ($F(2, 28) = 0.66, p = .542, \eta_p^2 = .045$).

5.2.3.3 Revisionszeiten

Deskriptiv zeigt sich, dass die Vpn bei einer niedrig vertrauenswürdigen Quelle der LO-Prämisse 3.52s ($SD = 1.99$ s) für ihre Revisionen benötigten. Wenn die Quelle einer mittleren Vertrauenswürdigkeit entsprach, benötigten die Vpn 4.59s ($SD = 2.41$ s) für die Revision. Bei einer hoch vertrauenswürdigen Quelle benötigten die Vpn 4.80s ($SD = 2.68$ s). Die deskriptiven Daten sind zur Veranschaulichung in Abbildung 16 abgetragen.

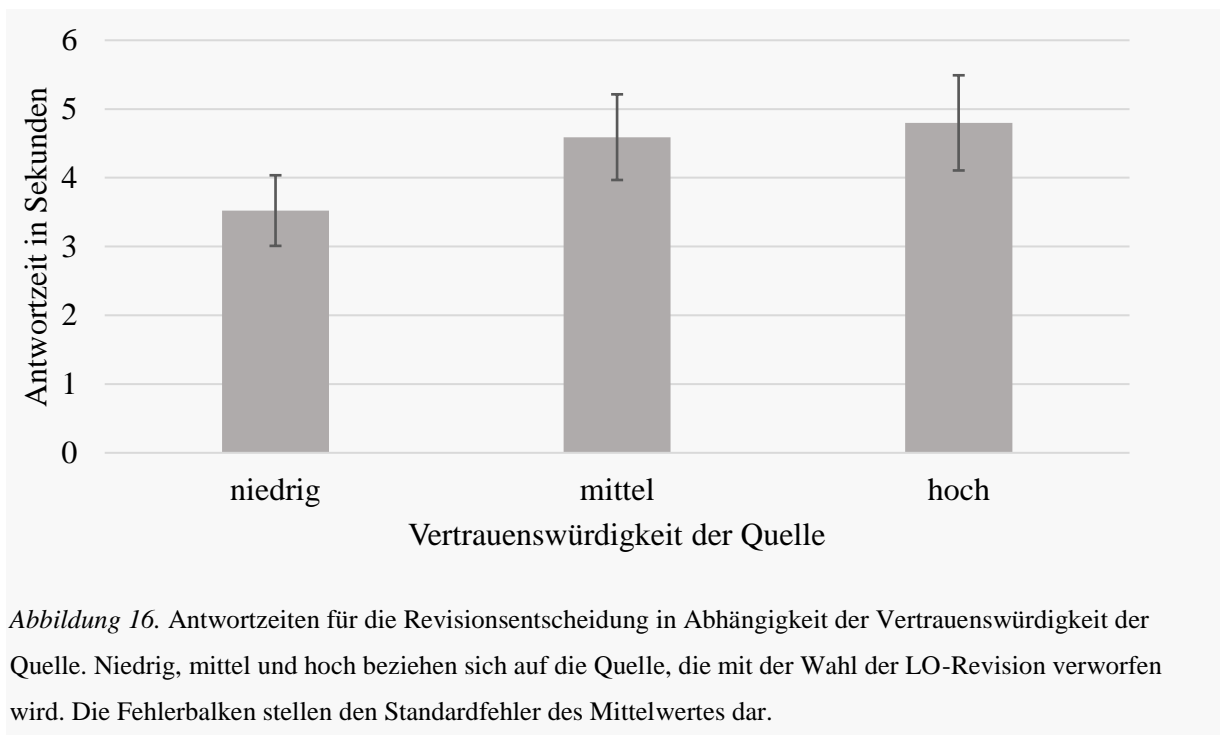


Abbildung 16. Antwortzeiten für die Revisionsentscheidung in Abhängigkeit der Vertrauenswürdigkeit der Quelle. Niedrig, mittel und hoch beziehen sich auf die Quelle, die mit der Wahl der LO-Revision verworfen wird. Die Fehlerbalken stellen den Standardfehler des Mittelwertes dar.

Um die deskriptiven Unterschiede inferenzstatistisch zu prüfen, habe ich eine einfaktorielle Varianzanalyse mit Messwiederholung für den Innersubjektfaktor „Vertrauenswürdigkeit“ (niedrig \times mittel \times hoch) durchgeführt. Die Analyse ergab einen signifikanten Haupteffekt für den Faktor „Vertrauenswürdigkeit“ ($F(2, 28) = 3.93, p = .031, \eta_p^2 = .219$). Um diesen Haupteffekt genauer zu erklären, habe ich drei post hoc t -Tests für verbundene Stichproben durchgeführt. Um dabei einer Alpha-Fehler Kumulation durch multiples Testen entgegenzuwirken, habe ich das Alpha-Niveau nach Bonferroni korrigiert (korrigiertes $\alpha = .017$). Die Vpn benötigten bei einer niedrig vertrauenswürdigen Quelle signifikant weniger Zeit für die Revision im Vergleich zu einer hoch vertrauenswürdigen Quelle ($t(14) = 2.83, p = .013, d = .54$). Die anderen beiden Vergleiche konnten das nach Bonferroni korrigierte Alpha-Niveau nicht unterschreiten (niedrig vs. mittel: $t(14) = 2.38, p = .032, d = .48$; mittel vs. hoch: $t(14) = 0.37, p = .714, d = .08$).

5.2.4 Diskussion

In dem vorliegenden Experiment habe ich das Quellenprinzip bei der räumlichen Überzeugungsrevision geprüft. Personen haben bei einer niedrig vertrauenswürdigen Quelle das LO-Prinzip genauso häufig verwendet wie bei einer hoch vertrauenswürdigen Quelle. Deshalb konnte mein Experiment meine zweite Hypothese, dass *das Quellenprinzip die Revisionsstrategie beim Revidieren räumlicher mentaler Modelle beeinflusst* nicht stützen.

Obwohl die Vpn die Berufe eindeutig als unterschiedlich vertrauenswürdig empfunden und gleichzeitig diese Berufe eindeutig im Hauptexperiment verarbeitet hatten, konnte die Häufigkeit der Revision nach dem LO-Prinzip nicht beeinflusst werden. Trotz dieses Befundes konnte ich zeigen, dass die Vpn ihre Entscheidungen nicht ganz unabhängig von dem Quellenprinzip trafen. Wenn die Vpn durch das LO-Prinzip die Prämisse einer niedrig vertrauenswürdigen Quelle verworfen hatten, benötigten sie 1.7 Sekunden weniger Zeit für ihre Revisionen, als wenn sie die Prämisse einer hoch vertrauenswürdigen Quelle verworfen. Dies deutet darauf hin, dass es den Personen leichter gefallen ist, die Information eines wenig vertrauenswürdigen Berufes zu verwerfen. Diese Befunde stimmen mit denen von Wolf et al. (2012) zum Quellenprinzip bei konditionalen Schlüssen überein. In ihrem ersten Experiment haben die Vpn ebenfalls eine Sekunde weniger Zeit benötigt, wenn sie die Informationen einer niedrig vertrauenswürdigen Quelle verworfen haben. Interessant ist dabei jedoch, dass die Vpn in ihrem Experiment deutlich länger (11.59s) für ihre Revisionen benötigten, wenn es keinen Unterschied in den Vertrauenswürdigkeiten gab (mittel/mittel Paarung). Die Autoren schlossen daraus, dass ein klarer Kontrast in der Vertrauenswürdigkeit das Denken bei der Revision erleichtert. In meinem Experiment benötigten die Vpn bei mittel vertrauenswürdigen Quellen nicht mehr Zeit für die Revisionen. Dieser auffällige Unterschied in den Befunden von Wolf et al. (2012) und meinem Experiment kann durch die unterschiedlichen Revisionsstrategien erklärt werden. In dem Experiment von Wolf et al. (2012) haben die Autoren konditionale Schlussfolgerungen als Basis für die Überzeugungsrevisionen verwendet. Sie zeigten, dass das Quellenprinzip das *mismatch principle* nach Johnson-Laird et al. (2004) überschreiben kann. Wenn die Quellen jedoch beide mittel vertrauenswürdig waren, glichen die Personen die mentalen Modelle nach dem *mismatch principle* ab. Es ist anzunehmen, dass dieser Abgleich kognitiv aufwendiger ist, als die Revision lediglich nach der Vertrauenswürdigkeit der Quelle zu richten. Bei relational räumlichen Schlussfolgerungen wie in meinem Experiment gleichen die Personen jedoch nicht die Modelle ab, sondern variieren das ursprüngliche Modell mit dem LO-Prinzip. Deshalb nehme ich an, dass ohne einen klaren Unterschied in der Vertrauenswürdigkeit die Revision kognitiv nicht aufwendiger ist und die Vpn deshalb im Vergleich zu Revisionen bei konditionalen Inhalten nicht mehr Zeit benötigten.

Eine wichtige methodische Kritik an dem vorliegenden Experiment ist jedoch die fehlende inhaltliche Relevanz der Prämissen zu den verwendeten Berufen. Alle Prämissen beschrieben die räumliche Relation von Objekten aus den Kategorien Obst oder Werkzeug. Die Quellen der Prämissen waren aber sehr unterschiedliche Berufe wie beispielweise ein

Polizist und ein Versicherungsvertreter. Deshalb ist es naheliegend, dass die Quellen – trotz eindeutiger vorheriger Verarbeitung – nicht genug Relevanz für die Revisionsentscheidungen hatten und der Effekt der Vertrauenswürdigkeit sich nicht eindeutiger nachweisen ließ. Aus diesem Grund habe ich ein weiteres Experiment durchgeführt und Prämissen verwendet, deren Inhalt auf die Berufe abgestimmt war. Um die Inhalte möglichst gut für die Berufe abzubilden, habe ich dazu keine räumliche, sondern andere Relationen (z. B. „Der rote Unfallwagen ist mehr wert als der blaue Unfallwagen“ für die Berufe Versicherungsvertreter und Polizist) verwendet. Knauff et al. (1998a) sowie Knauff und Johnson-Laird (2002) konnten dabei zeigen, dass auch nicht räumliche Relationen räumlich verarbeitet werden und ich habe die Vpn explizit instruiert, sich alle resultierenden Anordnungen auf einer horizontalen Achse (hoch → niedrig) vorzustellen. So konnte ich sicherstellen, dass die Aufgaben bis auf die Inhalte analog zum vorliegenden Experiment mit nur räumlichen Relationen verarbeitet wurden. Alle Prämissenpaare wurden in einer Vorstudie analog zu dem Vorgehen aus Wolf et al. (2012) ohne die Präsentation der Berufe auf Gleichwertigkeit getestet. So konnte sichergestellt werden, dass die Revisionsstrategie nicht ausschließlich durch die Inhalte der Prämissen beeinflusst wurde.

Die Ergebnisse der Revisionsentscheidungen zeigten deskriptiv, dass das LO-Prinzip am häufigsten gewählt wurde, wenn die zu verwerfende Prämisse von einer niedrig vertrauenswürdigen Quelle gegeben wurde ($M = 81.25\%$, $SD = 22.78\%$). Für mittlere ($M = 64.58\%$, $SD = 31.00\%$) und hoch ($M = 68.75\%$, $SD = 34.66\%$) vertrauenswürdige Quellen zeigten sich das LO-Prinzip vergleichbar häufig. Es konnte sich jedoch kein statistisch robuster Haupteffekt zeigen ($p = .095$). Auch für die Antwortzeiten ließ sich eindeutiger Quelleneffekt zeigen ($p = .821$). Das bedeutet, dass auch durch Anpassen der Inhalte kein eindeutiger Quelleneffekt auf die Revisionsentscheidungen nachgewiesen werden konnte.

Zusammenfassend lässt sich also sagen, dass sich das LO-Prinzip – wie bereits bei vielen vorherigen Untersuchungen (z. B. Bucher, Krumnack, Nejasmic, & Knauff, 2011; Bucher & Nejasmic, 2012; Knauff et al., 2013) – als die dominante Revisionsstrategie bei der räumlichen Überzeugungsrevision herausstellt. Personen haben sich nicht nach der Vertrauenswürdigkeit der Quelle, sondern nach der funktionellen Asymmetrie der Terme bei ihren Revisionsentscheidungen gerichtet. Bei räumlich abstrakten Prämissen fiel es den Vpn jedoch leichter, die Informationen einer wenig vertrauenswürdigen Quelle zu verwerfen. Dieser Befund gibt einen Hinweis darauf, dass soziale Faktoren die Überzeugungsrevision beeinflussen können. Mein Experiment und das nachträglich durchgeführte Kontrollexperiment legen nahe, dass besonders die wenig vertrauenswürdigen Quellen für die

Revisionsentscheidungen eine Rolle spielen können. Diese Befunde sind aber nicht eindeutig und müssen in weiteren Experimenten geprüft werden. Hier wäre es wichtig, realistischere Settings zu bilden und die Vertrauenswürdigkeiten in echten Gruppen zu manipulieren und zu prüfen, in wie weit das Revidieren von Überzeugungen durch wenig vertrauenswürdige Quellen erleichtert oder beeinflusst werden kann.

5.3 Zusammenfassung Kapitel 5

In diesem Kapitel habe ich untersucht, ob die räumliche Überzeugungsrevision durch soziale Faktoren wie die Konformität oder die Vertrauenswürdigkeit der Quelle beeinflusst werden kann. Experiment 6 zeigte, dass sich Personen deutlich durch die Gruppenmeinung beeinflussen ließen. Die Vpn entschieden sich viel häufiger für die LO-Revision, wenn der Rest einer fiktiven Gruppe sich ebenfalls für diese Revision entschieden hatte. Wenn die meisten Personen jedoch gegen diese Revision waren, entschieden sich die Vpn ebenfalls gegen das LO-Prinzip. Wenn nur eine oder zwei Personen von allen vier Personen die LO-Revision wählte, dann entschieden sich die Vpn zufällig für eine der beiden Revisionsalternativen. Dies ist ein Hinweis darauf, dass das Lösen der Aufgaben in einer (fiktiven) Gruppe die Vpn dazu veranlasste, die Revisionsalternativen genauer zu betrachten. Dies könnte dazu geführt haben, dass die Probanden die logische Gleichwertigkeit der LO- bzw. RO-Revision häufiger erkannten und deshalb ihre Revision eher nach dem *mismatch principle* (Johnson-Laird et al., 2004) wählten. Dass das Arbeiten in einer Gruppe die Leistung des einzelnen beeinflusst, ist als *soziale Erleichterung* (*social facilitation*, Zajonc, 1965) bekannt und konnte in neueren Studien auch durch die Anwesenheit von virtuellen Agenten beim Lösen mathematischer Probleme nachgewiesen werden (Karacora, Dehghani, Krämer-Mertens, & Gratch, 2012).

In Experiment 7 konnte kein direkter Einfluss der Vertrauenswürdigkeit auf die Revisionsentscheidungen nachgewiesen werden. Die Personen verwendeten unabhängig von dem Quellenprinzip gleichermaßen häufig das LO-Prinzip für ihre Revisionen. Trotzdem konnte ich zeigen, dass es den Vpn kognitiv leichter fällt, die Information einer niedrig vertrauenswürdigen Quelle aufzugeben. Dieser Befund hat zusammen mit den Ergebnissen aus meinem Experiment zur Konformität weitreichende Folgen für das räumliche Schlussfolgern in Gruppen. Wenn man das Denken von Gruppen untersuchen möchte, müssen optimale Bedingungen geschaffen werden, um den Konformitätsdruck einer Gruppe auf den Einzelnen zu minimieren. Es scheint den Personen leichter zu fallen, wenn sie für ihre Entscheidungen soziale Faktoren wie die Glaubwürdigkeit einer Quelle berücksichtigen

können. Die Vertrauenswürdigkeit von Personen kann jedoch neben ihrem Beruf über viele weitere soziale Dimensionen definiert werden. So kann man eine Person auch als vertrauenswürdiger erachten, wenn man dieser Person eine höhere Kompetenz für die zu lösende Aufgabe zuspricht (McCroskey & Teven, 1999). Für das Denken in Gruppen bedeutet dies, dass Konformitätseffekte unter Umständen reduziert werden können, wenn jedes Gruppenmitglied ausreichende Informationen zu den anderen Gruppenmitgliedern erhält. Dazu muss man sicherstellen, dass die Gruppenmitglieder davon überzeugt sind, dass alle Personen gleichermaßen fähig sind, die Probleme zu lösen. Soziale Faktoren sind also für das Denken in Gruppen sehr wichtig und sollten auf methodischer, aber auch auf theoretischer Ebene berücksichtigt werden.

Kapitel 6: Allgemeine Diskussion

Lag Dr. Watson mit seiner Aussage richtig? Ist die Gruppe ein Wetzstein für das menschliche Denken? Um diese Frage zu beantworten, habe ich das räumliche Schlussfolgern von Gruppen mit Hilfe der Phasen der Konstruktion, Variation und Revision mentaler Modelle getestet. Ich konnte zeigen, dass Personen in einer Gruppe sehr gut gemeinsam räumliche mentale Modelle von Problemen konstruieren und daraus Konklusionen ablesen können. Obwohl die Gruppen beim Konstruieren dem Kontinuitätseffekt unterlagen, waren sie in der Lage, mehr korrekte Modelle zu konstruieren als Individuen. Außerdem habe ich gezeigt, dass die Teilnehmer einer Gruppe versuchen, möglichst viele Informationen mit ihrem Partner zu tauschen, indem sie mentale Annotationen über die Diskontinuität der Prämissen im Gespräch berücksichtigen und diese ihrem Partner mitteilen. Bei der kritischen Phase des Denkens – der Variation mentaler Modelle – waren Gruppen allgemein den Individuen überlegen. Dieser Vorteil war auf deskriptiver Ebene bei den schwierigen indeterminierten Problemen am größten. Gruppen konnten sogar die vom *truth wins* Modell vorhergesagte Leistung übertreffen. Das aus meinen Daten erstellte *collaboration wins* Modell sagt vorher, dass Gruppen über das *truth wins* Modell hinaus ein großes Potential aufweisen. So konnten Gruppen bei den schwierigen indeterminierten Problemen selbst dann eine valide Konklusion ziehen, wenn ihre Gruppenmitglieder alleine dazu nicht in der Lage waren. Wenn Personen gezwungen waren, ihre mentalen Modelle zu revidieren, waren sie bereit, die ansonsten sehr robuste LO-Revisionsstrategie aufzugeben, um sich der Gruppenmeinung anzuschließen. Dieser Konformitätseffekt entstand dabei in beide Richtungen: Wenn die Majorität sich für das LO-Prinzip entschied, dann entschieden sich die Vpn ebenfalls häufiger für diese Revisionsalternative. Wenn die Majorität gegen die LO-Revision war, dann entschieden sich auch die Vpn gegen diese Revision. Aber auch bei weniger eindeutigen Präferenzen haben sich die Vpn beeinflussen lassen. Wenn nämlich die anderen Personen der fiktiven Gruppe sich nicht eindeutig für oder gegen die LO-Revision entschieden, dann wählten die Versuchspersonen scheinbar zufällig eine der beiden Alternativen. Obwohl das Soziale offensichtlich die Revision räumlicher Überzeugungen beeinflusst, war die Vertrauenswürdigkeit der Quelle von Informationen kaum von Bedeutung. Die Vertrauenswürdigkeit hatte keinen Einfluss darauf, wie häufig die Vpn die LO-Revision wählten. Das Quellenprinzip wirkte eher indirekt auf die Revision: Personen fiel es leichter, die Aussage einer wenig vertrauenswürdigen Quelle zu verwerfen.

Insgesamt zeigen meine Befunde, dass die Gruppe den Individuen beim Denken überlegen ist und dass soziale Faktoren einen Einfluss auf das Verarbeiten räumlicher

mentaler Modelle haben. Wir könnten also schlussfolgern, dass Dr. Watson richtig mit seiner Aussage lag: Gruppen beeinflussen das Denken. Somit könnten wir den Fall schließen und zu den Akten legen. Doch welcher Fall wird einfach so beendet, ohne die Hintergründe genau aufzuklären? Wie genau funktioniert das Denken in Gruppen? Und worauf ist die Überlegenheit der Gruppe zurückzuführen? Diese Fragen werde ich im nachfolgenden Abschnitt beantworten, indem ich die Befunde aus meinen Experimenten zusammen mit den bekannten Theorien zur Informationsverarbeitung in eine *Theorie kollaborativen räumlichen Schlussfolgerns* integriere.

Zunächst werde ich das wichtigste Axiom meiner Theorie erläutern. Dann zeige ich, wie das Denken in der Gruppe nach dieser Theorie erklärt wird und erläutere anschließend, welche Prozesse für das Denken in Gruppen wichtig sind. Danach gehe ich spezieller auf die Konsequenzen meiner Befunde für verschiedene Bereiche der Psychologie ein. Dabei diskutiere ich meine Befunde zunächst in Hinblick auf die Denkpsychologie und erörtere, ob andere Theorien des Denkens ebenfalls meine Befunde erklären können. Anschließend ordne ich meine Befunde in die sozialpsychologische Forschung ein und leite dann gesellschaftliche Konsequenzen aus meinen Befunden ab. Am Ende dieses Kapitels werde ich zeigen, welche Fragen noch unbeantwortet sind und wie diese in zukünftigen Experimenten geklärt werden könnten.

6.1 Kollaboratives räumliches Schlussfolgern - Eine Theorie des Denkens in Gruppen

Grundsätzlich nehme ich für die Theorie kollaborativen räumlichen Schlussfolgerns (KRS Theorie) an, dass das Denken in Gruppen durch ein Modell der Informationsverarbeitung abgebildet werden kann (siehe Hinsz et al., 1997). Die Informationsverarbeitung auf individueller Ebene besteht aus dem Input, der Verarbeitung und dem Output (Atkinson & Shiffrin, 1968). Für das räumliche Schlussfolgern bedeutet dies, dass die Prämissen den Input darstellen. Die Verarbeitung findet durch die Konstruktion, Inspektion und Variation mentaler Modelle statt (Knauff et al., 1998b). Der Output stellt die von den Individuen gezogene Konklusion dar. Ich nehme an, dass diese Art der Verarbeitung für die Gruppe sehr ähnlich ist. Dabei habe ich jedoch ein wichtiges Axiom, das der Theorie zu Grunde liegt: *Gruppen denken nicht*. Menschen repräsentieren Informationen mental und manipulieren diese Repräsentationen, um weitere Informationen abzuleiten. Die Gruppe ist dabei für mich kein eigenständig denkendes System. Vielmehr wirkt sie auf das Denken des Einzelnen und kann somit dazu führen, dass sich die Leistung der Gruppe qualitativ von der individuellen

Leistung unterscheidet. Deshalb bezieht sich meine Aussage „Gruppen ziehen mehr Konklusionen als Individuen“ auf den gesamten Prozess von der Verarbeitung der Prämissen bis zu einer – kollaborativ generierten – Antwort. Das Ziehen einer Konklusion ist jedoch ein ausschließlich individueller Prozess. Das bedeutet, dass die Gruppe keinen eigenständig denkenden Agenten darstellt und ihre Antwort von den Konklusionen ihrer Individuen supervenient wird (List & Pettit, 2006, 2011). Von Cranach, Oxsenbein und Valach (1986) schlagen für Gruppen zwei Formen der Informationsverarbeitung vor: die primäre und die sekundäre Informationsverarbeitung. Die primäre Verarbeitung findet auf Ebene der Individuen statt und umfasst alle kognitiven Prozesse. Diese kognitiven Prozesse formen jeweils die Basis für die sekundäre Ebene, die Gruppenebene. Hier werden aufgabenrelevante Informationen kommuniziert, auf die individuelle Ebene zurückgeführt und verarbeitet. Von Cranach et al. (1986) konzentrierten sich jedoch überwiegend auf handlungsrelevante Informationsverarbeitungen von Gruppen (z. B. das Führen eines Segelbootes). Ich nutze die Annahmen der Autoren und wende sie auf die grundlegenden kognitiven Prozesse des räumlichen Schlussfolgerns an. Das bedeutet, dass ich das Denken in Gruppen in zwei sich wechselseitig beeinflussenden Ebenen beschreibe: die individuelle Ebene und die Gruppenebene. Zunächst werde ich den gesamten Aufbau der KRS Theorie und das Denken in Gruppen mit beiden Ebenen beschreiben. Anschließend gehe ich spezifischer auf einzelne Aspekte dieser Prozesse ein und zeige, wie die Gruppe zu mehr validen Konklusionen kommt, aber auch, wie Fehler entstehen können.

6.1.1 Aufbau der KRS Theorie

Abbildung 17 stellt schematisch den Ablauf nach der KRS Theorie dar. Zunächst werden die Prämissen auf individueller Ebene verarbeitet. Das bedeutet, dass jedes Gruppenmitglied individuell eine integrierte Repräsentation – ein präferiertes mentales Modell – der Prämissen konstruiert. Falls die Prämissen mehr als eine Interpretation zulassen, halten die Personen eine Annotation über diese Interpretationen im Arbeitsgedächtnis. Wenn die Personen die Annotation berücksichtigen, dann variieren sie das präferierte mentale Modell und versuchen weitere Konklusionen zu finden. Im nächsten Schritt beginnt der eigentliche Gruppenprozess: die Deliberation. Jedes individuell konstruierte Modell und die vorläufigen Konklusionen dienen als Input für den Deliberationsprozess auf der Gruppenebene. Die Personen teilen ihre individuellen mentalen Modelle und ihre vorläufigen Konklusionen der Gruppe mit. Dabei gleichen die Individuen das von den anderen Gruppenmitgliedern geäußerte Modell mit dem eigenen Modell ab. Wenn alle Gruppenmitglieder das gleiche Modell konstruiert und die

gleiche Konklusion gezogen haben, kann die Gruppe direkt zum Output übergehen und eine Antwort tätigen. Wenn die Personen Inkonsistenzen entdecken, dann müssen diese Inkonsistenzen diskutiert und aufgelöst werden. Ziel der Gruppe ist es dabei, so viele Informationen wie möglich aus allen individuellen Modellen zu erhalten. Einige Personen sind jedoch gezwungen, ihr ursprüngliches Modell aufzugeben und zu revidieren. Am Ende dieses Deliberationsprozesses entsteht eine gemeinsame Repräsentation des Problems: ein geteiltes mentales Modell. Dieses geteilte mentale Modell wird wieder auf individueller Ebene encodiert und inspiziert. Jedes Individuum liest darauf eine Konklusion aus dem geteilten mentalen Modell ab und bespricht diese mit der Gruppe. Wenn sich die Gruppe einig ist, können sie eine Antwort geben. Wenn die Konklusionen jedoch Inkonsistenzen aufweisen, dient jedes individuell encodierte mentale Modell wieder als Input für einen neuen Deliberationsprozess. Dieser Prozess wird solange wiederholt, bis eine konsistente Konklusion von allen Mitgliedern aus dem geteilten mentalen Modell gezogen werden kann. Wichtig bei dieser Theorie ist, dass die Prozesse nicht notwendigerweise strikt seriell ablaufen. So ist es möglich, dass eine Person bereits das geteilte mentale Modell encodiert hat, während ein anderes Gruppenmitglied noch sein ursprüngliches Modell revidiert. Das heißt, dass das Denken in Gruppen deutlich dynamischer zu verstehen ist, als dies durch Abbildung 17 suggeriert wird.

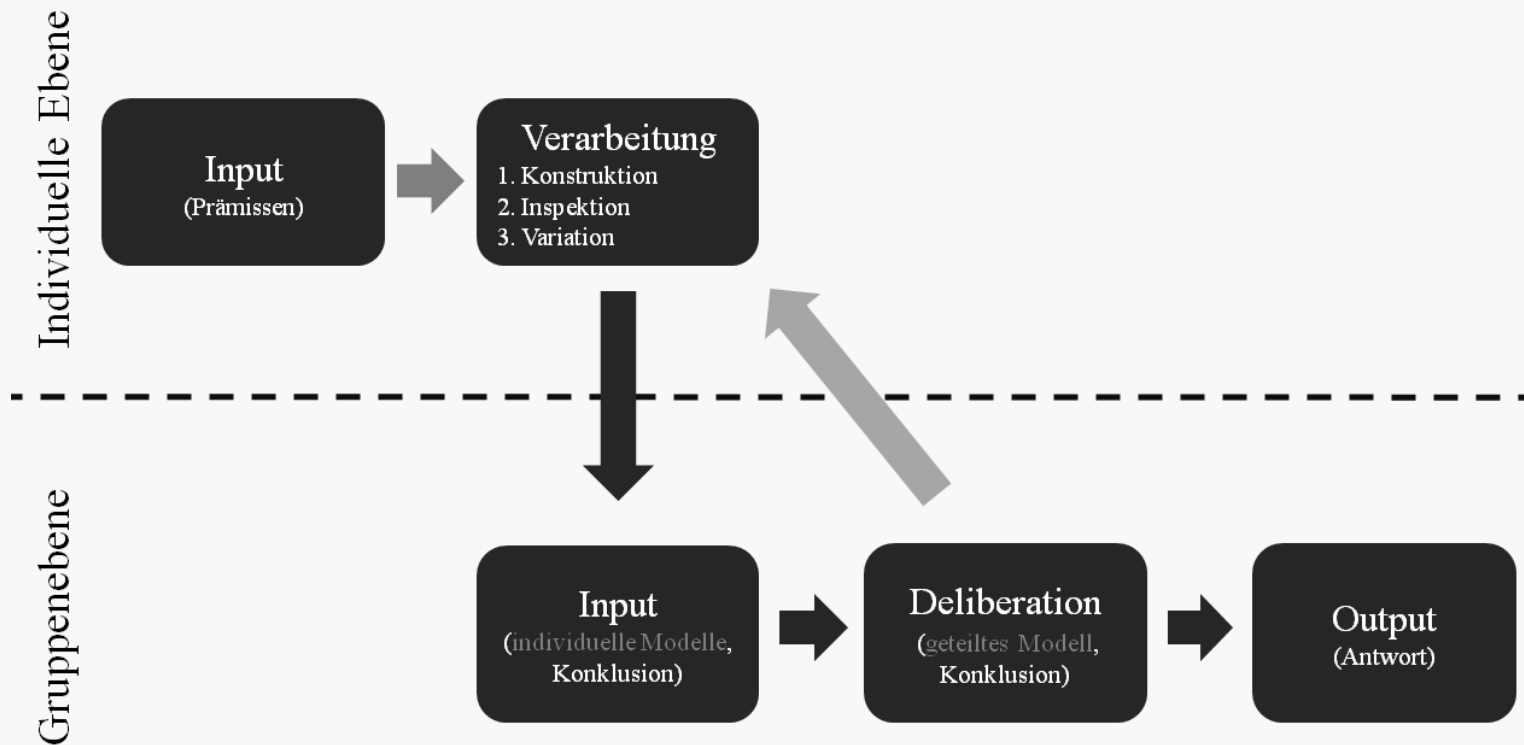


Abbildung 17. Ablauf des kollaborativen räumlichen Schlussfolgerns. Die Gruppenmitglieder verwenden ihre individuellen Modelle als Input für den Deliberationsprozess. Im Deliberationsprozess werden die individuellen mentalen Modelle und Konklusionen abgeglichen. Wenn alle Gruppenmitglieder das gleiche Modell konstruiert und die gleiche Konklusion gezogen haben, kann die Gruppe direkt zur Antwort übergehen. Ansonsten muss ein neues geteiltes mentales Modell aus den individuellen Modellen konstruiert werden. Dieses geteilte Modell wird dann von jedem Gruppenmitglied encodiert. Jede Person liest darauf individuell eine Konklusion ab. Diese Konklusionen dienen wieder als Input für den nächsten Deliberationsprozess. Wenn die Konklusionen nicht einheitlich sind, werden die Modelle wieder geprüft. Ansonsten wird die geteilte Konklusion als Antwort gegeben.

Das Denken auf Gruppenebene läuft also in drei Schritten ab. (1) Die zuvor individuell konstruierten mentalen Modelle und Konklusionen dienen dem Deliberationsprozess als Input. (2) Die Gruppe prüft die individuellen Modelle auf Unterschiede und konstruiert gegebenenfalls ein neues geteiltes mentales Modell. (3) Die Gruppenmitglieder lesen eine Konklusion ab und geben eine gemeinsame Gruppenantwort. Wenn sie im dritten Schritt feststellen, dass sie zu unterschiedlichen Konklusionen kommen, wiederholen sie den ersten Schritt um zu überprüfen, ob sie das geteilte mentale Modell unterschiedlich encodiert haben.

Die primäre Annahme der KRS Theorie ist also, dass das Denken ein ausschließlich individueller Prozess ist. Deshalb unterliegt die Gruppe dem aus den Untersuchungen zum individuellen Denken bekannten Kontinuitätseffekt (Kapitel 3). Doch wenn das Denken der Gruppe nur auf individuellem Denken basiert, warum konnten die Gruppen dann in allen drei Experimenten zum Unbestimmtheitseffekt (Kapitel 4) mehr valide Konklusionen ziehen als Individuen? Und wie kann es sein, dass mein *collaboration wins* Modell sogar bei den schwierigen indeterminierten Problemen eine so gute Gruppenleistung vorhersagt? Bereits beim individuellen Denken konnten die indeterminierten Probleme uns helfen, das menschliche Denken zu verstehen. Und auch beim kollaborativen räumlichen Schlussfolgern geben diese Probleme am meisten Aufschluss darüber, wie die Gruppe als epistemischer Wetzstein für das Denken funktionieren kann.

6.1.2 Die Gruppe als epistemischer Wetzstein

Ich nehme an, dass während des Deliberationsprozesses zwei wichtige Faktoren der Gruppe das Potential bieten, das individuelle Denken zu übertreffen. Der erste Faktor stützt sich auf die Annahme, dass die Gruppe mehr Informationen berücksichtigen kann und deshalb in der Lage ist, ein kohärenteres geteiltes mentales Modell eines Problems zu konstruieren (Hinsz et al., 1997; Maciejovsky & Budescu, 2007). Mit mehr Informationen meine ich jedoch nicht, dass die Gruppe mehr Prämissen verarbeitet. Allen Personen standen in meinen Experimenten exakt die gleichen Informationen zur Verfügung und auch die Individuen waren häufig in der Lage, korrekte Schlussfolgerungen aus den Prämissen zu ziehen. Vielmehr basiert das geteilte mentale Modell der Gruppe auf mehreren individuellen Modellen. Die Gruppe versucht, die Variabilität der individuellen mentalen Modelle zu reduzieren und strebt dabei Konvergenz an (Kolfshoten & Brazier, 2013). Deshalb versuchen sie, möglichst viele Informationen aus den individuellen Modellen zu integrieren und ein kohärentes geteiltes mentales Modell des Problems zu konstruieren. Zusätzlich sucht die Gruppe nach anderen Modellen, bei denen die

Konklusion nach dem geteilten mentalen Modell nicht hält. Wenn ein Gruppenmitglied nicht in der Lage war, eine korrekte Repräsentation zu konstruieren, versuchen die anderen Gruppenmitglieder diese Fehler zu korrigieren und die Person dazu zu bewegen, ihre ursprüngliche Überzeugung zu revidieren. Die Revision der individuellen mentalen Modelle richtet sich dabei auf der einen Seite nach der Majorität der Gruppenmeinung. Das heißt, dass das Modell gewählt wird, welches von den meisten Gruppenmitgliedern konstruiert wurde. Auf der anderen Seite versuchen die Gruppenmitglieder sich gegenseitig von dem richtigen mentalen Modell zu überzeugen. Als *richtiges Modell* ist die korrekte Integration der Prämissen in das präferierte Modell, sowie – falls notwendig – das Berücksichtigen einer Annotation über weitere mögliche Modelle zu verstehen. Am Ende entsteht durch den Deliberationsprozess eine bessere Basis (das geteilte mentale Modell) für das Denken. Deshalb führt das Denken in der Gruppe zu mehr validen Konklusionen im Vergleich zum individuellen Denken. Allein dieser Faktor reicht aber nicht aus, um die Vorhersagen des *collaboration wins* Modells zu erklären. Würde ausschließlich die Integration der individuellen Modelle verantwortlich für die Gruppenleistung sein, dann sollte im Falle, dass alle Gruppenmitglieder die Prämissen falsch in ein Modell integrieren oder die Annotation nicht berücksichtigen, das geteilte mentale Modell ebenfalls zu einer unvaliden Gruppenkonklusion führen. Dies zeigte sich jedoch nicht in meinen Experimenten. Deshalb muss während des Deliberationsprozesses ein weiterer kognitiver Prozess für die Überlegenheit der Gruppe verantwortlich sein: die Metakognition.

Unter Metakognition versteht man das Wissen und Kognitionen über die eigenen kognitiven Prozesse (Flavell, 1979). Das bedeutet, dass man seine eigenen Denkprozesse überwachen kann und falls es durch die Aufgabe erforderlich ist, diese Denkprozesse den Anforderungen anpasst. Die Metakognition ist in ihren Ursprüngen eine Theorie aus der Lern- und Gedächtnispsychologie (für einen Überblick: Koriat, 2007). Dabei prägte Hart (1965) für das Abrufen von Gedächtnisinhalten den Begriff des *feeling of knowledge* (FOK). Das FOK ist nach Hart eine Intuition darüber, dass man eine Antwort auf eine Frage weiß, sie aber nicht abrufen kann. Thompson, Turner und Pennycook (2011) haben für das schlussfolgernde Denken einen ganz ähnlichen metakognitiven Prozess untersucht. Sie haben in einer ganzen Reihe von Experimenten gezeigt, dass Menschen beim Ziehen einer Konklusion ein *feeling of rightness* (FOR) entwickeln. Dieses FOR ist eine Art Bauchgefühl (Thompson, 2009) darüber, dass man eine folgerichtige Konklusion gezogen hat. Wenn man sich sehr sicher ist, eine richtige Konklusion gezogen zu haben, dann empfindet man ein hohes FOR. Wenn man jedoch eine hohe Unsicherheit über die Richtigkeit der Konklusion verspürt, dann empfindet

man ein niedriges FOR. Thompson et al. (2011) zeigten, dass ein hohes FOR mit deutlich schnelleren Antworten einherging und die Konklusion in einer zweiten Runde deutlich seltener hinterfragt wurde. War jedoch das FOR niedrig, haben die Personen ihre ursprüngliche Konklusion häufiger hinterfragt. Gab man ihnen die Möglichkeit, ihre ursprüngliche Konklusion zu revidieren, haben Personen mit einem niedrigen FOR dies häufiger getan als Personen, die ein hohes FOR empfanden. Ich nehme an, dass genau dieses FOR eine entscheidende Rolle dabei spielt, dass Gruppen den Individuen beim Denken mit räumlichen Relationen überlegen sind.

Meine Annahme ist, dass das Interagieren in Gruppen zu einer Erhöhung der metakognitiven Prozesse führt. Das bedeutet, dass die Gruppenmitglieder ihre eigenen Denkprozesse stärker überwachen und somit häufiger eine Unsicherheit über ihre Konklusionen verspüren. Wenn ein Gruppenmitglied ein präferiertes mentales Modell konstruiert hat und die Annotation über weitere Modelle nicht im Arbeitsgedächtnis halten konnte, dann entwickelt das Mitglied ein Gefühl darüber, dass seine Konklusion möglicherweise falsch ist. Das Gruppenmitglied entwickelt also ein geringes FOR zu seiner Konklusion. Dieses FOR wird stärker berücksichtigt, wenn die Person weiß, dass sie ihre Konklusion gemeinsam mit anderen Gruppenmitgliedern besprechen wird. Wenn die Personen also ein geringes FOR für ihre Konklusion haben, äußern sie diese Unsicherheit im Deliberationsprozess. Dann kann jedes Gruppenmitglied prüfen, ob es ebenfalls dieses Gefühl teilt oder eventuell sogar die Annotation im Arbeitsgedächtnis gehalten hat. Das bedeutet, dass durch den Deliberationsprozess in der Gruppe die Annotation bei indeterminierten Problemen häufiger berücksichtigt wird. Das Interagieren in Gruppen führt also dazu, dass wir unsere eigenen Denkprozesse genauer überwachen. Deshalb kann eine Gruppe mehr valide Konklusionen ziehen als Individuen.

Dass Menschen beim Schlussfolgern ein Gefühl über die Richtigkeit ihrer Konklusionen entwickeln, konnte bereits in vielen Untersuchungen gezeigt werden (für einen Überblick siehe: De Neys, 2014). Dabei zeigte sich dieses Gefühl beim Denken mit Syllogismen (Klauer & Singmann, 2013) und konnte dabei sogar durch physiologische Maße wie die Hautleitfähigkeit (De Neys, Moyens, & Vansteenwegen, 2010) oder Blickbewegungen (Ball, Phillips, Wade, & Quayle, 2006) nachgewiesen werden. Für den Einfluss einer Gruppeninteraktion auf das FOR der Gruppenmitglieder gibt es bisher nur wenige Evidenzen. Fisher, Knobe, Strickland und Keil (2017) konnten jedoch zeigen, dass das Kooperieren in Gruppen dazu führt, dass die Gruppenmitglieder die Richtigkeit ihrer Annahmen häufiger hinterfragen. Weitere Hinweise für die Wichtigkeit des FOR konnten

Bahrami et al. (2010) bei einfachen visuellen Entscheidungsaufgaben nachweisen. Ihre Befunde legen nahe, dass gerade das Äußern einer subjektiven Sicherheit beim Interagieren von Dyaden zu mehr korrekten Antworten führt. So war nicht einmal Feedback zu den (korrekten) Lösungen der Aufgaben notwendig. Wenn die Personen ihre subjektive Sicherheit über zuvor gegebene individuelle Antworten in der Dyade frei äußern und diskutieren durften, konnte die Dyade deutlich mehr korrekte Antworten finden als Individuen. Aufbauend auf diesen Befunden führte Koriat (2012) sehr ähnliche Experimente durch und zeigte, dass nicht einmal eine direkte Interaktion der Dyade notwendig ist. Vielmehr reichte es aus, wenn Personen die subjektive Sicherheit einer anderen Person ohne direkte Interaktion erhielten und diese für ihre Antworten nutzen konnten. Er zeigte jedoch auch, dass Gruppen bei besonderen Fragen schlechter als Individuen abschnitten. Diese Fragen zeichnen sich dadurch aus, dass sie grundsätzlich oft falsch beantwortet werden und die Personen trotzdem eine hohe subjektive Sicherheit zu der Richtigkeit ihrer Antwort empfinden (z. B. „Ist die Hauptstadt von Australien Sydney oder Canberra?“).

Im Bereich der *hidden profiles* (Stasser, 1988) postulierten Scholten, Knippenberg, Nijstad und De Dreu (2007) ein zum FOR sehr ähnliches Konstrukt: die *epistemische Motivation*. Unter epistemischer Motivation in Gruppen verstehen die Autoren den Wunsch eines Gruppenmitglieds, ein tieferes Verständnis des Entscheidungsprozesses der Gruppe zu entwickeln. Die Autoren zeigten, dass Gruppen, deren Mitglieder eine hohe epistemische Motivation hatten, mehr Informationen untereinander austauschten und somit zu besseren Entscheidungen gelangten. Die epistemische Motivation bezieht sich jedoch überwiegend auf den Austausch von Informationen in Gruppen. Das heißt, dass die epistemische Motivation sich vor allem in der Überwachung und Steuerung des Gruppenprozesses zeigt. Wohingegen das FOR als Metakognition eine Überwachung der individuellen Prozesse darstellt. Dennoch zeigt das Konstrukt der epistemischen Motivation, dass es unterschiedliche Neigungen beim Denken in Gruppen geben kann und diese die Gruppenleistung beeinflussen.

Auch das *collaboration wins* Modell ist ein Hinweis auf eine höhere Berücksichtigung des FOR in Gruppen. Nach diesem Modell können Dyaden sogar dann eine valide Konklusion ziehen, wenn beide Individuen zuvor alleine nicht die richtige Konklusion finden konnten. Das bedeutet, dass die Dyaden in der Lage waren, die Annotation über weitere Interpretationen der Prämissen zu berücksichtigen und eine valide Konklusion zu ziehen. Ich nehme an, dass das FOR dazu führte, dass die Gruppenmitglieder ihre Unsicherheit über weitere mögliche Modelle geäußert haben und die Dyade deshalb die Annotation schließlich doch berücksichtigen konnte. Deshalb beinhaltet das *collaboration wins* Modell nicht nur die

Wahrscheinlichkeit, dass eine Person das andere Gruppenmitglied von dem korrekten mentalen Modell überzeugen kann. Vielmehr enthält es auch die Wahrscheinlichkeit, dass eine Person mit niedrigem FOR diese Unsicherheit im Deliberationsprozess äußert und die Gruppe somit dennoch die richtige Lösung finden kann. Diese Annahme wird nicht nur durch mein *collaboration wins* Modell gestützt. In meinem Experiment 2 konnte ich bei der Konstruktion mentaler Modelle zeigen, dass Personen, die gemeinsam mit einem Partner in einer Dyade ein mentales Modell konstruierten, die Annotation von diskontinuierlichen Problemen häufiger verbalisierten als Personen, die ausschließlich alleine arbeiteten. Deshalb denke ich, dass die Gruppe ähnlich zu dem Effekt der sozialen Erleichterung (*social facilitation*; Zajonc, 1965) die metakognitiven Prozesse der Überwachung und Steuerung verstärkt und somit das Denken jedes Gruppenmitglieds schärft. Das Denken in Gruppen dient also als „epistemischer Wetzstein“ für den Geist.

6.1.3 Fehler beim kollaborativen räumlichen Schlussfolgern

Bisher habe ich beschrieben, weshalb die Gruppe den Individuen beim Denken überlegen ist. Doch wie entstehen Fehler beim kollaborativen Denken? Ich nehme an, dass dies vor allem auf den Integrationsprozess der individuellen mentalen Modelle in ein geteiltes mentales Modell zurückzuführen ist. Wenn ein Gruppenmitglied ein präferiertes mentales Modell konstruiert hat und die Annotation über weitere Möglichkeiten vernachlässigt, sollte sie ein niedriges FOR entwickeln und dieses in der Gruppe äußern. Doch was passiert, wenn alle Gruppenmitglieder die Annotation vernachlässigen? Ich nehme an, dass dies zu einer Akzentuierung der Denkprozesse (Hinsz et al., 1997) führt. Das bedeutet, dass die Gruppenmitglieder sich voreilig auf die Konklusion nach dem präferierten mentalen Modell einigen und selbst metakognitive Prozesse wie ein niedriges FOR nicht ausreichen, um die Annotation zu berücksichtigen. Dabei kann das präferierte Modell wie eine von den Gruppenmitgliedern geteilte Information aus dem *hidden profile* Paradigma nach Stasser und Titus (1985) verstanden werden. Wir wissen aus den Untersuchungen zum *hidden profile*, dass Gruppenmitglieder besonders die geteilten Informationen diskutieren und andere wichtige Informationen für ihre Antworten vernachlässigen (Schulz-Hardt et al., 2006; Stasser & Abele, 2019). Ich nehme an, dass dies in sehr ähnlicher Weise auch auf die präferierten mentalen Modelle der Gruppenmitglieder zutrifft. Das präferierte mentale Modell führt im Sinne eines *gemeinsamen Wissens Effekts* (*common knowledge effect*; Gigone, 2010) dazu, dass keine weiteren Modelle berücksichtigt werden.

Aufgrund der insgesamt sehr wenigen Fehler in meinen Experimenten zum Unbestimmtheitseffekt konnte ich keine statistisch robuste Analyse dieser Annahme durchführen. Meine deskriptiven Daten zeigen jedoch, dass die Fehler der Dyaden bei indeterminierten Problemen in fast allen Fällen auf ein korrekt konstruiertes Modell zurückzuführen sind. Dies deutet darauf hin, dass die Dyaden die Annotation über ein zweites Modell vernachlässigt haben. Bei Individuen zeigte sich ein etwas anderes Muster: Fehler waren deskriptiv seltener auf ein korrekt konstruiertes Modell zurückzuführen. Das bedeutet, dass Individuen häufig nicht einmal ein korrektes Modell konstruiert hatten und deshalb zufällig antworteten. Dieses Muster kann auch durch die deskriptiven Befunde aus der Studie von Roberts (2000) gestützt werden. Dieser zeigte, dass Fehler von Individuen bei indeterminierten Problemen sogar häufig auf rotierte Versionen der korrekten Modelle zurückzuführen sind. Sie bieten jedoch keine explizite Analyse über die präferierten mentalen Modelle, sodass auch ihre Rückschlüsse nicht gesichert sind. Insgesamt weisen aber meine deskriptiven Befunde und jene von Roberts (2000) darauf hin, dass Fehler von Individuen und Dyaden nicht exakt dem gleichen Muster folgen.

6.2 Konsequenzen für die Denkpsychologie

Die Erforschung menschlicher Denkprozesse hat eine lange Tradition. Dies ist nicht überraschend, wenn man bedenkt, wie allgegenwärtig das Schlussfolgern für den Menschen ist. Und trotzdem ist es nicht einfach, die zu Grunde liegenden Denkprozesse zu erklären. Ein großer Beitrag zur Klärung dieser Prozesse konnte im Bereich des räumlichen Schlussfolgerns geleistet werden (Johnson-Laird & Byrne, 1991; Knauff, 2006, 2013; für einen Überblick siehe Kießner & Ragni, 2017). Diese Forschungen basieren jedoch ausschließlich auf den Denkprozessen von Individuen. Meine Befunde sind deshalb ein Novum in der Erforschung des räumlichen Schlussfolgerns. Ich zeige, dass eine individualistische Sicht nicht ausreicht, um das gesamte Spektrum menschlichen Denkens erklären zu können. Menschen sind soziale Wesen (Dunbar, 1998; Mercier & Sperber, 2017; Sloman & Fernbach, 2018) und dies trifft auch auf das Denken zu. Menschen können in Gruppen nicht nur bessere Repräsentationen eines räumlichen Problems bilden, sondern auch mehr valide Konklusionen ziehen. Außerdem ließen sich die Personen selbst beim individuellen Denken durch soziale Faktoren wie die Konformität beeinflussen. Doch welche Konsequenz haben meine Befunde für die denkpsychologische Forschung? Bedeutet dies, dass das menschliche Denken ausschließlich im Gruppenkontext untersucht werden sollte?

Dies wäre vermutlich ein wenig zu weit gegriffen. Ohne die Forschung an einzelnen Personen hätten wir niemals so viel über die Denkprozesse des Menschen lernen können. Aber es ist dennoch wichtig, die soziale Komponente des Denkens nicht zu unterschätzen. Dabei ist dieser Gedanke gar nicht so neu. Die *Theorie sozialer Verträge* (*social contract theory*; Cosmides, 1989) besagt, dass der Mensch das Schlussfolgern entwickelt hat, um gute und schlechte Intentionen anderer Personen zu identifizieren. Das bedeutet, dass die Vertreter dieser Theorie eine evolutionäre Funktion des Denkens vermuten. Menschen prüfen die Aussagen anderer und versuchen zu schlussfolgern, ob man es möglicherweise mit einem Betrüger zu tun hat (Cosmides, Tooby, Fiddick, & Bryant, 2005). Tatsächlich konnte man zeigen, dass die *Wason Selection Task* (Wason, 1966) deutlich häufiger korrekt gelöst werden kann, wenn man sie in den Kontext sozialer Verträge einbettet (z. B. Wenn du X für mich machst, dann werde ich Y für dich tun) (Cosmides, 1989; Gigerenzer & Hug, 1992). Der größte Nachteil dieser Theorie ist jedoch, dass sie eine modulare Theorie ist (Cosmides & Tooby, 2000; Gigerenzer & Hug, 1992). Das bedeutet, dass ihre Vorhersagen lediglich auf das Denken mit sozialen Inhalten zutreffen und kaum etwas über allgemeine Denkprinzipien oder das Denken in Gruppen aussagen können. Meine Befunde und die KRS Theorie sind nicht modular und leisten deshalb einen wichtigen Beitrag für die Entwicklung einer allgemeinen Theorie zum Denken in Gruppen.

6.2.1 Andere Denkprobleme

Obwohl meine Befunde und die daraus abgeleitete KRS Theorie auf den Experimenten zum räumlichen Schlussfolgern basieren, können wir von ihnen auch einiges für die anderen Domänen des Denkens lernen. Beim konditionalen Schlussfolgern führen Moshman und Geil (1998) die Überlegenheit der Gruppe auf den Diskussionsprozess zurück. Sie erläutern jedoch nicht genau, was in der Diskussion dazu führt, dass die Gruppe mehr valide Konklusionen ziehen kann. Beim Denken mit Syllogismen schließen Tindale et al. (2012) aus ihren Befunden, dass die Gruppe vor allem durch das Teilen der Repräsentationen den Individuen überlegen sein können. In der KRS Theorie mache ich ganz ähnliche Annahmen und füge den wichtigen Prozess der Metakognition als Erklärung für das kollaborative Denken hinzu. Die Metakognition als Erklärung für Denkprozesse ist keine neue Idee (Thompson et al., 2011). Was jedoch bisher nicht untersucht wurde, ist der Fakt, dass die Gruppe als sozialer Motor für die Metakognition funktionieren kann. In meinen Experimenten nehme ich an, dass das Denken in der Gruppe dazu führt, dass die Annotation häufiger berücksichtigt wird. Bei der *Wason Selection Task* gibt es keine Indeterminiertheit des Problems. Aber dennoch müssen

auch hier mehrere Möglichkeiten über eine präferierte Antwort hinaus berücksichtigt werden (Johnson-Laird & Wason, 1970). Dies könnte erklären, weshalb die Befunde von Moshman und Geil (1998) deskriptiv sehr ähnlich zu dem von mir gefundenen *collaboration wins* Modell sind. Gruppen konnten die Lösung auch dann finden, wenn ihre Mitglieder alleine nicht dazu in der Lage waren. Deshalb vermute ich, dass das *collaboration wins* Modell ebenfalls für das kollaborative Denken mit anderen Problemen, wie der *Wason Selection Task*, eine Erklärung bieten kann. Ich nehme an, dass auch beim Denken mit konditionalen oder Syllogismen die Gruppenmitglieder durch die Deliberation mehr Gewicht auf ihr subjektives Empfinden der Richtigkeit (FOR) der Konklusion legen. So sollte das FOR dazu führen, dass sie auf der einen Seite ihre Überlegungen häufiger diskutieren und auf der anderen Seite mehr individuelle Informationen in ein geteiltes mentales Modell der Probleme integrieren.

6.2.2 Alternative Theorien zum Denken

Ich habe sowohl für die Vorhersage meiner Experimente als auch für die KRS Theorie die an Individuen gut untersuchte Theorie mentaler Modelle (TMM) als Grundlage verwendet. Auch wenn beim räumlichen Schlussfolgern wenig Dissens darüber herrscht, dass die TMM das individuelle Denken am besten beschreiben kann (Knauff & Knoblich, 2017), sollten meine Befunde nicht isoliert von anderen Theorien interpretiert werden. Deshalb werde ich zunächst meine Befunde in Hinblick auf die größte Gegentheorie – die Theorie mentaler Logik – und einer relativ neuen Strömung – die probabilistische Theorie – diskutieren und prüfen, ob diese Theorien das räumliche Schlussfolgern in Gruppen besser erklären können als die TMM.

6.2.2.1 Theorie mentaler Logik

Gruppeninteraktionen basieren häufig auf sprachlicher Kommunikation und Argumentation (Mercier & Sperber, 2011). Deshalb ist es naheliegend anzunehmen, dass das Schlussfolgern mit räumlichen Inhalten von Gruppen besser durch eine mehr an Sprache orientierter Theorie erklärt werden kann. Die Theorie mentaler Logik (TML) nach Rips (1994) und die syntaktischen Theorien für das räumliche Schlussfolgern von Hagert und Hansson (1983) und Hagert (1985) machen genau diese Annahmen. Anstelle von modellhaften Repräsentationen halten die Personen die Prämissen in ihrer syntaktischen Struktur im Arbeitsgedächtnis und manipulieren diese, indem sie verschiedene Regeln anwenden. Nachdem sie die Konklusionen ähnlich einer mathematischen Formel folgend abgeleitet haben, werden diese syntaktischen Informationen in ein sprachliches Format zurückgeführt. Gruppenmitglieder müssen fortwährend ihre mentale Repräsentation mit den anderen Mitgliedern kommunizieren

und teilen. Man könnte annehmen, dass es für die Gruppenmitglieder viel leichter ist, eine sprachähnliche Repräsentation im Arbeitsgedächtnis zu halten und zu verbalisieren. Im besten Fall könnten die Personen die Prämissen sprachnaher Form im Arbeitsgedächtnis halten und im Deliberationsprozess wiederholen und besprechen, bis die Gruppe gemeinsam aus den Prämissen eine Konklusion abgeleitet hat. Das bedeutet, dass man annehmen könnte, dass die Theorie mentaler Logik das räumliche Schlussfolgern in der Gruppe besser beschreiben kann, als die KRS Theorie mit ihrer zu Grunde liegenden TMM.

Diese Annahme kann jedoch durch meine Befunde zum Kontinuitätseffekt nicht gestützt werden. In meinem Experiment 1 konnte ich zeigen, dass Dyaden dem gleichen Kontinuitätseffekt unterliegen, der bereits in mehreren Experimenten zum individuellen Denken gefunden wurde (Ehrlich & Johnson-Laird, 1982; Knauff et al., 1998a; Nejasmic et al., 2015; Oakhill & Garnham, 1985; Sugimoto und Kusumi, 2014). Der Kontinuitätseffekt basiert auf der Annahme, dass die Prämissen als ein zusammenhängendes mentales Modell repräsentiert werden (Nejasmic et al., 2015). Dementsprechend sollte sich der Kontinuitätseffekt nicht zeigen, wenn man die Prämissen in sprachnaher Form repräsentiert (Ehrlich & Johnson-Laird, 1982). Meine Ergebnisse sprechen aber nicht für diese Annahme. In Experiment 2 zur kollaborativen Konstruktion mentaler Modelle habe ich getestet, in welcher Form Gruppenmitglieder zwei zuvor repräsentierte Prämissen mit ihrem Partner teilen. Dabei konnte ich auf der einen Seite zeigen, dass Personen in einer Dyade in 81 % der Fälle ihre Repräsentation in Form von Modellen (z. B. „Apfel Birne Mango Kiwi“ oder „Apfel Birne und Mango Kiwi“) verbalisiert haben. Auf der anderen Seite haben die Vpn ihrem Partner in nur 19 % der Fälle den exakten Wortlaut der Prämissen mitgeteilt. Deshalb sprechen meine Befunde zum Kontinuitätseffekt gegen die Annahme, dass die Vpn in einer Gruppe die Prämissen als sprachnahe Propositionen repräsentiert und verarbeitet haben.

Die Befunde aus meinem Kapitel 4 zum Unbestimmtheitseffekt sind auf den ersten Blick weniger eindeutig. Byrne und Johnson-Laird (1989) haben die indeterminierten Probleme konstruiert, um die Vorhersagen der TML gegen die der TMM zu testen. Nach der TML sollten indeterminierte Probleme nicht schwieriger sein als determinierte. Die TMM nimmt jedoch an, dass indeterminierte Probleme deutlich schwieriger sind, da mehr als ein Modell berücksichtigt werden muss. Dieses Muster konnte ich für die Dyaden bei keinem der drei Experimente finden. Indeterminierte Probleme waren in den meisten Fällen genauso schwierig wie determinierte Probleme. Ist es also möglich, dass das kollaborative Schlussfolgern beim Unbestimmtheitseffekt besser durch die TML als durch die TMM beschrieben werden kann?

Die Antwort auf diese Frage liegt zum einen in meiner KRS Theorie. Zum anderen kann mein Aufgabenmaterial uns einen Hinweis darauf geben, ob die TML das Denken besser beschreiben kann. Nach der KRS Theorie sollten Gruppen besonders bei den schwierigen indeterminierten Problemen profitieren. Personen, die in Gruppen gemeinsam denken, berücksichtigen ihr FOR durch den sozialen Kontext häufiger. Das heißt, dass sie die Annotation bei indeterminierten Problemen mit den Gruppenmitgliedern besprechen und mehr valide Konklusionen ziehen. Für zwei-Modelle determinierte Probleme reicht es aus, ein einziges korrektes Modell zu konstruieren und die Konklusion abzulesen. Das heißt, dass keine Annotation über weitere Modelle notwendig ist und ein geringes FOR seltener entsteht. Ein geringes FOR ist nach der KRS Theorie jedoch ausschlaggebend dafür, dass die Probleme intensiver durchdacht und in der Gruppe besprochen werden. Deshalb sollte der Vorteil des Denkens in Gruppen nach der KRS Theorie am stärksten bei indeterminierten Problemen ausfallen. Das heißt, dass sich die Leistung der Gruppe bei indeterminierten Problemen an die Leistung von determinierten Problemen annähern sollte und die Probleme deshalb vergleichbar schwierig sind. Da ich aber das FOR in keinem Experiment explizit geprüft habe, ist diese Annahme aber noch sehr theoretisch orientiert. Deshalb benötigen wir zusätzliche Hinweise, um zu prüfen, ob die TML das Denken in der Dyade besser abbildet als die TMM.

Eine genauere Antwort können uns interessanterweise die ein-Modell Probleme in meinen Experimenten geben. Ich habe als Grundlage für meine Experimente die Probleme aus Roberts (2000) Studie zum Unbestimmtheitseffekt verwendet. Dieser wollte prüfen, ob verschiedene Präsentationsarten dazu führen, dass die Probanden die TMM oder TML im Sinne einer kognitiven Strategie unterschiedlich bevorzugen. Die Probleme waren so konstruiert, dass ein-Modell Probleme nach der TML von Rips (1994) und der Erweiterung für räumliches Schlussfolgern nach Hagert und Hansson (1983) sowie Hagert (1985) schwieriger sein sollten als zwei-Modelle determinierte Probleme. Doch wie genau sind die Vorhersagen zu verstehen? Bei allen Problemen in meinen Experimenten zum Unbestimmtheitseffekt habe ich immer nach der Relation zwischen den Termen D und E gefragt. Meine ein-Modell und zwei-Modelle determinierten Probleme waren so konstruiert, dass die Terme E und D immer mit den Termen A und C verbunden waren. Um die Relation zwischen A und C herauszufinden, mussten bei ein-Modell Problemen die ersten beiden Prämissen (z. B. „A ist links von B“ und „C ist rechts von B“) berücksichtigt werden. Bei den zwei-Modelle determinierten Problemen war es jedoch immer so, dass die Relation von A und C durch eine einzelne Prämisse explizit genannt wurde (z. B. „A ist links von C“). Somit

sollten die Konklusionen bei zwei-Modelle determinierten Problemen nach der TML syntaktisch leichter ableitbar sein, als bei ein-Modell Problemen. Dieses Muster konnte ich jedoch in keinem meiner Experimente nachweisen. Für alle Vpn – Individuen und Dyaden – waren ein-Modell Probleme die am leichtesten zu lösenden Probleme.

Zusammenfassend lässt sich also sagen, dass es sehr unwahrscheinlich ist, dass das Denken in der Gruppe am ehesten durch die TML beschrieben wird. Anhand meiner Befunde zum Kontinuitätseffekt und Unbestimmtheitseffekt können wir annehmen, dass die TMM mit der Ergänzung des metakognitiven Prozesses des FOR das Denken in Gruppen am besten erklären kann.

6.2.2.2 Probabilistische Theorie

In den letzten Jahren gewann für das individuelle Denken eine neue theoretische Perspektive immer mehr an Popularität. Nach diesem *neuen Paradigma der Denkpsychologie* (Over, 2009) müssen Prämissen und eine daraus gezogene Konklusion nicht unbedingt wahr oder falsch sein. Das bedeutet, dass diese Theorien im Gegensatz zu der TMM und TML den Anspruch der Monotonie für das Schlussfolgern aufgeben (Oaksford & Chater, 1995). Das Schlussfolgern ist in diesem Paradigma also ein Schließen mit Unsicherheiten und man kann eine Konklusion nur mit einer bestimmten Wahrscheinlichkeit als wahr oder falsch akzeptieren (Oaksford, 2015). Deshalb nennt man dieses Paradigma auch die *probabilistische Theorie* (PT) des Denkens (Oaksford & Chater, 2019). Um Unsicherheiten zu lösen, nutzen Menschen ihr Hintergrundwissen (De Neys, Schaeken, & D’Ydewalle, 2003), ihre Emotionen (Perham & Oaksford, 2005) oder eigene Präferenzen (Bonnefon, Girotto, & Legrenzi, 2012). Dabei ist das Denken nach der PT häufig in ein Konversationssetting eingebettet (Oaksford & Chater, 2019). Das heißt, dass jemand schlussfolgert, um sich selbst auf eine Diskussion vorzubereiten und sinnvolle Argumente für die eigene Überzeugung abzuleiten (Mercier & Sperber, 2011). Diese Argumente können eine unterschiedliche Stärke haben und deshalb unterschiedlich überzeugend sein (Hahn & Oaksford, 2007). In meiner KRS Theorie nehme ich ebenfalls an, dass die Gruppenmitglieder sich gegenseitig beim Schlussfolgern von den korrekten Repräsentationen überzeugen. Bedeutet dies also, dass die PT das Denken in der Gruppe besser erklären kann als die TMM?

Bei der Beantwortung dieser Frage scheitert die PT beim räumlichen Schlussfolgern genau an ihrem wichtigsten Argument: dem Hintergrundwissen. Die Aufgaben meiner Experimente waren so konstruiert, dass sie möglichst kein Hintergrundwissen möglich machten. So ist es kaum denkbar, einen Grund zu finden, warum ein Apfel nicht links von

einer Birne sein sollte. Deshalb ist es sehr unwahrscheinlich, dass Individuen oder Dyaden den Prämissen in meinen Problemen eine gewisse Wahrscheinlichkeit zugeordnet haben und daraus ihre Konklusionen gezogen haben. Für das Schlussfolgern mit räumlichen Relationen kann die PT das Denken also nicht beschreiben. Aber was kann sie zu der Revision räumlicher Überzeugungen sagen?

Ich konnte in meinem Experiment 6 zeigen, dass die Revision durch Konformität beeinflusst werden kann. In diesem Fall ist es möglich, dass die Probanden einer bestimmten Revisionsalternative (LO oder RO) durch den Gruppeneinfluss unterschiedliche Wahrscheinlichkeiten zugesprochen haben. Das bedeutet nach der PT, dass sie sich für die Alternative entscheiden, die sie auf Grund der Gruppenmeinung für wahrscheinlicher erachten. Dies könnte auch erklären, warum die Vpn in dem Experiment bei nicht eindeutigen Gruppenmeinungen (eine oder zwei Personen) sich eher zufällig für die LO- bzw. RO-Revision entschieden haben. Wenn die Möglichkeiten gleich wahrscheinlich wirkten und der Rest der Gruppe sich auch nicht eindeutig entscheiden konnte, dann haben die Vpn ebenfalls eher zufällig eine Alternative gewählt. Es ist also möglich, dass die PT die Ergebnisse für mein Experiment zur Konformität erklären kann. Wenn es aber darum geht, die genauen Prozesse bei der Revision zu erklären, stößt die PT ebenfalls an ihre Grenzen. Die PT beschreibt das Denken nur auf computationaler Ebene (Marr, 1982). Das bedeutet, dass sie uns sagen kann, welchen Zweck das Denken hat. Sie kann uns sagen, dass einer bestimmten Revision (z. B. durch den Einfluss der Gruppe) weniger Wahrscheinlichkeit zugesprochen und diese deshalb seltener gewählt wurde. Die Theorie kann aber nichts darüber sagen, wie genau der Prozess der Revision funktioniert. Das heißt, dass die PT nichts über die algorithmische Beschreibungsebene (Marr, 1982) aussagen kann. Dies kann ich ebenfalls an den Befunden aus meinem Experiment 7 verdeutlichen. In diesem Experiment habe ich das Quellenprinzip bei der Überzeugungsrevision untersucht. Nach der PT wird einer Prämisse weniger Wahrheit zugesprochen, wenn sie von einer niedrig vertrauenswürdigen Quelle geäußert wurde (Wolf et al., 2012). Das bedeutet, dass die Prämisse zu mehr Unsicherheit führt und deshalb eher verworfen wird. In meinem Experiment konnte ich jedoch keinen solchen Einfluss nachweisen und die PT schlägt keine alternative Revisionsstrategie vor. In meinem Experiment entschieden sich die Probanden in den meisten Fällen dazu, den Fakt unabhängig von der Quelle nach der LO-Revision zu variieren. Dieses Prinzip basiert jedoch auf der lokalen Revision eines ursprünglich konstruierten mentalen Modells (Knauff et al., 2013). Deshalb ist anzunehmen, dass meine Befunde in Experiment 7 deutlich besser durch die TMM als durch die PT erklärt werden können.

Zusammenfassend lässt sich sagen, dass die PT keine umfassende Erklärung für die Prozesse beim kollaborativen räumlichen Schlussfolgern bieten kann. Meine Befunde zur räumlichen Überzeugungsrevision kann sie zwar teilweise erklären, scheitert jedoch bei der algorithmischen Ebene des Denkens. Deshalb ist es für die Gruppenforschung, aber auch für die Erforschung der individuellen Überzeugungsrevision, weiterhin wichtig, Denktheorien wie die Theorie mentaler Modelle als Grundlage zu verwenden.

6.3 Konsequenzen für die Sozialpsychologie

Ivan Steiner stellte bereits vor einem halben Jahrhundert die berühmte Frage: „*Whatever happened to the group in social psychology?*“ (Steiner, 1974). Er kritisierte schon damals, dass die Gruppe als Forschungsobjekt für die Sozialpsychologie immer mehr an Bedeutung verlor. Den größten Grund dafür sah Steiner darin, dass die psychologische Forschung des Menschen immer mehr auf eine individualistische Ebene reduziert wurde. Dies wurde vor allem dadurch erklärt, dass die Gruppenforschung sehr zeitaufwendig ist und eine angemessene Stichprobengröße nur schwer realisierbar sei (Davis, 1996). Außerdem kritisierte Steiner (1974), dass es für die damalige Forschung keine ausreichende theoretische Basis für die Untersuchung von Gruppen gab. Erst mit dem Ansatz, die Gruppe als informationsverarbeitendes System zu betrachten, bekam die Gruppenforschung neuen Aufschwung (Brauner & Scholl, 2000). So behielt die Theorie der Informationsverarbeitung auf der einen Seite viel von der individualistischen Sicht auf die Gruppe bei. Es ermöglichte den Forschern aber auf der anderen Seite auch genauere Annahmen über die Gruppenprozesse zu tätigen und es konnte gezeigt werden, dass die Gruppe mehr Informationen verarbeitet (Laughlin et al., 2002), bessere Strategien verwendet (Laughlin et al., 2003) und eine geteilte Repräsentation eines Problems bildet (Maciejovsky & Budescu, 2007). Meine Befunde zum Denken in Gruppen können diese Evidenzen stützen. Aber zusätzlich kann die KRS Theorie der sozialpsychologischen Erforschung von Gruppen einen wichtigen Beitrag über diese Befunde hinaus liefern. Ein Problem an der bisherigen Erforschung von Gruppen ist, dass ihre Befunde häufig nur Aufschlüsse zur Gruppenleistung bei ganz speziellen Problemen geben können. Das bedeutet, dass Sozialpsychologen im Grunde nur sagen können, dass eine Gruppe bei einer ganz bestimmten Aufgabe (z. B. dem Buchstaben-zu-Zahlen Problem) überlegen ist. Die Befunde aus meiner Arbeit sollten jedoch einen deutlich allgemeineren Anspruch erfüllen können. Das Schlussfolgern stellt nicht lediglich ein spezielles Problem dar, welches es zu lösen gilt. Vielmehr ist es ein grundlegender epistemischer Prozess, den jeder Mensch fortwährend nutzt, um die Welt zu verstehen. Die TMM, die meinen Befunden

zu Grunde liegt, ist eine allgemeine Theorie des Denkens. Auch wenn die TMM in anderen Bereichen der Deduktion umstritten ist (Evans, Over, & Handley, 2005), können wir annehmen, dass die überlegene Gruppenleistung auch bei anderen Aufgaben auf die Konstruktion und Verarbeitung eines geteilten mentalen Modells zurückzuführen ist. Das gut untersuchte Buchstaben-zu-Zahlen Problem zeigt, dass es deutlich leichter gelöst werden kann, wenn man deduktive Schlüsse verwendet (Laughlin et al., 2002, 2003). Deshalb können wir annehmen, dass die überlegene Gruppenleistung auch bei diesem Problem auf Konklusionen aus mentalen Modellen der Aufgaben zurückzuführen ist.

Ein weiterer wichtiger Punkt ist, dass die sozialpsychologische Forschung sich häufig sehr auf den Anwendungsaspekt der Gruppenforschung konzentriert. So wird untersucht, wie Gruppen im Arbeitskontext kollaborieren (Walsh & Maloney, 2007) und Informationen austauschen (Greitemeyer & Schulz-Hardt, 2003). Mit meiner Arbeit zeige ich jedoch, dass das Denken als grundlegender kognitiver Prozess dabei nicht vernachlässigt werden sollte. Wenn wir mehr darüber lernen, wie Gruppen denken und durch welche Prozesse dieses Denken erklärt werden kann, dann können wir auch zuverlässige Vorhersagen und Methoden für die anwendungsorientierte Forschung entwickeln. Es wäre jedoch falsch, sich nur auf anwendungsorientiertes Problemlösen zu konzentrieren und die allgemeinen Denkprozesse des Schlussfolgerns zu vernachlässigen.

In der KRS Theorie postuliere ich, dass die Gruppe die Metakognition ihrer Gruppenmitglieder beeinflusst. Dass metakognitive Prozesse von Gruppen wichtig sein können, ist keine neue Annahme. Jedoch ist die Betrachtungsweise der Metakognition von Gruppen eine zu der KRS Theorie grundlegend andere. Die Metakognition von Gruppen kann insgesamt unter dem Begriff *social sharedness* zusammengefasst werden (Kameda, Tindale, & Davis, 2003; Tindale & Kameda, 2000). Auf der einen Seite wird angenommen, dass die Gruppenmitglieder nicht nur ausschließlich eine Repräsentation des Problems bilden, sondern vielmehr auch eine Repräsentation darüber konstruieren, welche Fähigkeiten die anderen Gruppenmitglieder besitzen und wie diese zur Lösung eines Problems beitragen können (*transaktives Gedächtnismodell*; Wegner, 1987). Auf der anderen Seite definiert sich die Metakognition der Gruppe als eine Repräsentation der Gruppenmitglieder darüber, wie Informationen in der Gruppe verteilt sind und wie diese Informationen verarbeitet werden (Hinsz, 2004; Tindale & Winget, 2019). Das bedeutet, dass die Theorien zur Metakognition von Gruppen die Vorstellung der Gruppenmitglieder über ihre eigenen Gruppenprozesse beschreiben. Die KRS Theorie sieht den Einfluss der Gruppe auf die individuelle Metakognition als Grund für die Überlegenheit der Gruppe. Trotzdem kann das Konzept der

social sharedness auch für die KRS Theorie bedeutsam sein. Wenn jemand gemeinsam mit anderen in einer Gruppe schlussfolgert, kann es hilfreich sein, eine Repräsentation über die Denkprozesse der Gruppe zu entwickeln. Wenn die Gruppenmitglieder wissen, dass die anderen gegebenenfalls die Annotation und weitere Interpretationen der Prämissen im Arbeitsgedächtnis halten sollten, dann können sie sich eher durch das eigene FOR dazu veranlasst sehen, diese Annotation in dem Deliberationsprozess zu diskutieren. Das heißt, dass man einerseits untersuchen sollte, inwieweit die *social sharedness* eine Rolle beim kollaborativen Denken spielt. Andererseits müssen die Einflüsse der Gruppe auf die individuelle Metakognition genauer untersucht werden und die sozialpsychologische Forschung kann dabei Vorhersagen aus der KRS Theorie ableiten. Es stellt sich außerdem die Frage, inwieweit eine Wechselwirkung zwischen der individuellen Metakognition beim Denken und der Metakognition der Gruppe besteht. Ich nehme an, dass – wie beim Schlussfolgern mit mentalen Modellen in der KRS Theorie – zunächst die individuelle (metakognitive) Repräsentation am bedeutsamsten ist und diese dann einen Einfluss auf die Metakognition über die Gruppenprozesse nehmen kann. So kann ich erst eine Repräsentation darüber bilden, wie die Gruppe zu einer folgerichtigen Konklusion gelangen kann, wenn ich selbst eine Repräsentation des Problems und meines Denkprozesses konstruiert habe.

Eine andere Perspektive auf die Gruppe als denkendes System haben Woolley, Chabris, Pentland, Hashmi und Malone (2010) eröffnet. Sie zeigten, dass man für eine Gruppe einen eigenen Intelligenzfaktor messen kann. Analog zu dem individuellen *g-Faktor* (*general intelligence*; Spearman, 1904) ermittelten die Autoren einen *c-Faktor* (*collective intelligence*), indem sie Gruppen verschiedene komplexe Aufgaben lösen ließen. Die Autoren zeigten, dass dieser *c-Faktor* zuverlässig die Leistung der Gruppe bei anderen kognitiven Aufgaben vorhersagen konnte. Das interessanteste Ergebnis ist jedoch, dass der *c-Faktor* der Gruppe nur in einem geringen Zusammenhang mit der Intelligenz der Gruppenmitglieder stand. Diese Befunde konnten in weiteren Untersuchungen gestützt werden (für eine Zusammenfassung: Woolley, Aggarwal, & Malone, 2015). Das bedeutet, dass diese Befunde im Gegensatz zu den Annahmen der KRS Theorie stehen. Das wichtigste Axiom der KRS Theorie ist, dass das Denken grundsätzlich ein individueller Prozess ist. Die Gruppe kann durch die Förderung metakognitiver Prozesse und das Integrieren individueller mentaler Modelle als epistemischer Wetzstein fungieren. Jedoch sollte nach der KRS Theorie die Denkfähigkeit des Einzelnen mit der Leistung der Gruppe im Zusammenhang stehen. Wie können diese Unterschiede erklärt werden? Dazu können uns die Aufgaben, die Woolley et al. (2010) für die Ermittlung des *c-Faktors* verwendet haben, einen Hinweis geben. So haben die Autoren unter anderem

Aufgaben zum Brainstorming, zum Generieren von Worten oder Schätzaufgaben verwendet. Obwohl Woolley et al. (2010) sogar Schlussfolgerungsaufgaben (Raven Matrizen Test; Raven, 1998) und räumliche Probleme (mentale Rotationen) verwendet haben, waren die Aufgaben so konzipiert, dass sie in die Kategorie der *additiven* oder *komplementäre Aufgaben* nach Steiner (1966, 1972) zugeordnet werden können. Das bedeutet, dass die Leistung der Gruppenmitglieder entweder addiert werden konnte oder dass die Gruppen die Aufgaben in kleinere Subtypen aufteilen konnten. Deshalb war die Gruppe in der Lage eine von der individuellen Intelligenz unabhängige Leistung in den Problemen zu erzielen. Die Probleme des kollaborativen Schlussfolgerns können jedoch am ehesten der disjunktiven Aufgabenstruktur nach Steiner (1966, 1972) zugeordnet werden. Die Gruppenmitglieder müssen jeweils ein vollständiges mentales Modell des Problems bilden und auf individueller Ebene eine Konklusion aus diesem Modell ablesen. Deshalb sollte der c-Faktor für das kollaborative Denken wenig prädiktiv sein. Dennoch wäre es interessant zu prüfen, inwieweit ein c-Faktor für disjunktive Aufgaben und besonders für das räumliche Schlussfolgern gefunden werden kann. Wenn sich ein solcher Faktor finden ließe, dann sollte dieser jedoch nicht unabhängig von der Intelligenz der Gruppenmitglieder sein und außerdem durch metakognitive Prozesse wie das FOR erklärt werden können.

Zusammenfassend lässt sich sagen, dass meine Befunde einen wichtigen Beitrag für die sozialpsychologische Forschung leisten können. So sollten Gruppenprozesse nicht nur bei anwendungsorientierten Problemen, sondern auch für grundlegendere Denkprobleme untersucht werden. Außerdem sollte in weiteren Experimenten der Einfluss der Gruppe auf die individuelle Metakognition jedes Gruppenmitglieds getestet werden. Dabei könnte geprüft werden, inwieweit die Metakognition der Gruppe und die individuelle Metakognition interagieren.

6.4 Konsequenzen für die Gesellschaft

Gruppen sind für den Menschen in vielen Situationen sehr bedeutsam und wir verlassen uns häufig auf die Gruppe bei wichtigen Entscheidungen (Hastie & Kameda, 2005). Doch unterliegen Gruppen auch Fehlern beim Denken und große Unglücke in der Menschheitsgeschichte können auf solche Fehler zurückgeführt werden (Janis, 2008; Reason, 1987). Ist es deshalb ratsam, sich weniger auf die Gruppe und dafür mehr auf Entscheidungen von Individuen zu verlassen? Der Großteil meiner Befunde zeigt, dass dies ein Fehler wäre. Gruppen können ein wichtiges Werkzeug sein. Selbst kleine Gruppen aus zwei Personen, wie in meinen Experimenten, konnten bessere Schlussfolgerungsleistungen erzielen als einzelne

Personen. Auch wenn in meinen Experimenten sich nicht immer ein starker Unterschied zwischen Individuen und Gruppen zeigte, ist es doch sehr oft ratsam, selbst kleine Vorteile der Gruppe zu nutzen. Bei Operationen interagieren Ärzte fortwährend miteinander und es zeigte sich, dass besonders räumliche Prozesse die Leistung der Ärzte vorhersagen können (Vajsbaher, Schultheis, & Francis, 2018; Mathias, Vogel, & Knauff, 2019). Dyaden konnten in meinen Untersuchungen beim räumlichen Schlussfolgern die individuelle Leistung übertreffen. Wenn also eine Person schwer krank ist und auf eine komplizierte Operation angewiesen ist, dann kann jede operative Leistungssteigerung – sei sie noch so gering – durch die Hinzunahme einer weiteren Person sehr bedeutend sein.

Schlussfolgernde Denkprozesse sind aber auch in der Schule sehr wichtig. So werden deduktive Aufgaben schon früh als didaktisches Werkzeug verwendet, damit Schüler Diskrepanzen in eigenen Versuchen entdecken und daraus ein besseres Verständnis der zu Grunde liegenden Prozesse entwickeln (Tröbst, Hardy, & Möller, 2011). In der aufwendigen *Längsschnittstudie zur Genese individueller Kompetenzen* (LOGIK; Schneider, 2008) haben Bullock und Sodian (2003) die Entwicklung von abstrakten und wissenschaftlichen Kompetenzen von der Kindheit bis in das Erwachsenenalter untersucht. Sie zeigten, dass bereits Grundschüler grundlegende logische Fähigkeiten besitzen und diese Fähigkeiten im Zusammenhang mit Kompetenzen wie das Hypothesentesten oder das Planen von Experimenten im Erwachsenenalter stehen. Daraus schließen sie, dass Kinder bereits früh in wissenschaftlichen und kritischen Denkprozessen ausgebildet und gefördert werden sollten. In experimentellen Untersuchungen zeigte sich jedoch, dass es Kindern im Grundschulalter oft sehr schwer fällt, mehr als ein mögliches mentales Modell für ihre Schlussfolgerungen zu berücksichtigen, wenn sie alleine arbeiten (Barrouillet, Gauffroy, & Lecas, 2008). Deshalb sollten Kinder deduktive Aufgaben gemeinsam in Gruppen lösen. So können die Kinder einerseits gemeinsam mehr über logische Denkaufgaben lernen. Andererseits können sich metakognitive Strategien beim Denken bereits früh entwickeln und durch die Kollaboration in Gruppen gefördert werden.

Das Denken in Gruppen ist also auch gesellschaftlich sehr bedeutsam. Ich habe jedoch auch gezeigt, dass starke kognitive Prinzipien wie die LO-Revision durch simple Manipulation der Konformität überschrieben werden können. Dies zeigt deutlich, wie wichtig es ist, sich Einflüssen wie der Konformität auf das Denken bewusst zu sein. Selbst wenn die Gruppe zu besseren Schlussfolgerungen gelangen kann, sollten wir auf experimenteller Ebene, aber auch in gesellschaftlicher Hinsicht diese Effekte nicht unterschätzen und versuchen, einen Weg zu finden, sie möglichst gering zu halten.

6.5 Weiterführende Forschung

Ich habe mit meinen Experimenten einige wichtige Aspekte beim Denken und Revidieren mit räumlichen Problemen untersucht. Wie in jeder Arbeit zum Denken ergeben sich jedoch noch weitere Fragen, die durch meine Arbeit nicht beantwortet werden konnten. Deshalb möchte ich im folgenden Teil auf weitere interessante Fragestellungen aufmerksam machen und Anregungen für zukünftige Forschungen geben.

6.5.1 Aufgabenschwierigkeit und weitere Effekte beim Denken

Ein wichtiger Aspekt, der in weiteren Experimenten untersucht werden sollte, ist die Vorhersage über das Entstehen von Fehlern nach der KRS Theorie. Ich nehme an, dass Fehler besonders dann entstehen, wenn alle Gruppenmitglieder zunächst das gleiche *präferierte mentale Modell* (Knauff et al., 1995) konstruieren und dieses in ein geteiltes präferiertes mentales Modell integrieren. Deskriptiv waren alle Fehler der Dyade in meinen Experimenten auf ein korrekt konstruiertes Modell zurückzuführen. Die Fehler der Individuen hingegen waren häufiger nicht so eindeutig. Insgesamt konnte ich durch die geringen Beobachtungen jedoch keine genauen Aussagen darüber treffen, ob dieser Trend robust ist. Deshalb müssen zukünftige Experimente nicht nur mit mehr Beobachtungen durchgeführt werden, sondern sollten auch schwierigere Denkaufgaben verwenden, um Fehler beim kollaborativen Denken zu evozieren.

Mit der Aufgabenschwierigkeit ist eine weitere sehr wichtige Frage zum Denken in Gruppen verbunden: Was geschieht, wenn die Probleme so schwierig sind, dass sie für ein Individuum kaum noch zu lösen sind? Kann die Gruppe dann trotzdem das individuelle Denken übertreffen? Die Schwierigkeit von räumlich relationalen Problemen kann über mehrere Faktoren beeinflusst werden. So könnte man beiden Problemen lediglich mehr Terme und Prädikate hinzufügen. Weitere Möglichkeiten die Schwierigkeit zu manipulieren sind die Komplexität der Relationen (Halford et al., 1998) oder die kognitiven Schritte, die für eine Konklusion notwendig sind (Ragni & Knauff, 2013). Ich nehme an, dass mit steigender Komplexität der Probleme, das FOR als metakognitiver Prozess an Bedeutung verliert. Vielmehr sollte die überlegende Denkleistung der Gruppe nun vor allem durch das geteilte mentale Modell und dessen Konstruktion aus den individuellen Modellen erklärt werden. Solange die Gruppenmitglieder die Möglichkeit haben, ihr mentales Modell zu demonstrieren und im Deliberationsprozess in ein geteiltes mentales Modell der Gruppe zu überführen, sollte die Gruppe das Potential haben, die individuelle Leistung beim Denken zu übertreffen. Wenn jedoch die Informationen die Arbeitsgedächtniskapazität der Gruppenmitglieder

überschreiten, sollte die Gruppe ebenfalls an ihre Grenzen stoßen. So wäre es zwar zunächst noch möglich, dass die einzelnen Gruppenmitglieder Teile des gesamten Modells encodieren und mit den anderen Gruppenmitgliedern austauschen. Um eine Konklusion ablesen zu können, muss jedoch eine integrierte Repräsentation des gesamten Problems konstruiert werden. Wenn die Anzahl der Terme oder die Komplexität der Relationen zu hoch werden, sollte das Denken in Gruppen deshalb ebenfalls scheitern. Dabei wäre interessant, in zukünftigen Experimenten zu prüfen, ob die kumulierte Arbeitsgedächtnisleistung der Gruppe linear von jedem zusätzlichen Gruppenmitglied abhängt, oder ob durch besondere Strategien der Gruppe (z. B. vorläufige Integration von Teilmodellen) mehr Kapazitäten für das Denken ermöglicht werden können. Andererseits ist anzunehmen, dass die Kommunikation der Modelle mit steigender Aufgabenschwierigkeit immer aufwendiger werden sollte. Deshalb sollte sich der *collaboration load* (Dillenbourg & Betrancourt, 2006) erhöhen und die Leistung der Gruppe beeinträchtigen.

Neben der Aufgabenschwierigkeit können wir auch durch andere Effekte beim Denken etwas über die Gruppe lernen. So zeigt der *Termreihenfolgeeffekt* (*figural effect*; Johnson-Laird & Bara, 1984), dass eine Konklusion nicht unabhängig von der Position der Terme in den Prämissen gezogen wird. Beim Termreihenfolgeeffekt werden die Terme (A, B, C) in unterschiedlicher Reihenfolge in den Prämissen genannt und es zeigt sich, dass – obwohl am Ende immer das gleiche Modell „ABC“ resultiert – diese Reihenfolgen zu unterschiedlichen Konklusionen führen können (Knauff et al., 1998a). Da ich annehme, dass auch das Denken in Gruppen vor allem ein individueller Prozess ist, sollten Gruppen einerseits auch diesen Effekt zeigen. Andererseits sind die Modelle, die nach der KRS Theorie als Input für den Deliberationsprozess dienen, bereits integrierte mentale Modelle der Prämissen. Deshalb könnte in einem Experiment zum kollaborativen Termreihenfolgeeffekt dieser Effekt für die Gruppe weniger Einfluss haben als für Individuen.

Zusätzlich zum Termreihenfolgeeffekt konnte gezeigt werden, dass selbst nicht räumliche Relationen wie soziale Vergleiche (besser/schlechter) oder andere Relationen wie „heller/dunkler“ räumlich repräsentiert und verarbeitet werden (De Soto, London, & Handel, 1965). Dabei zeigte sich für die Relation „klüger als“ ein *Inferenzeffekt* (Potts, 1972). Dieser besagt, dass die Probanden eine Konklusion (z. B. ArD), die sich aus den Prämissen (ArB, ArC, CrD; Modell: ABCD) ergibt, besser und schneller erinnern konnten, als lediglich eine der genannten Prämisse (z. B. ArB) (Potts, 1972). Außerdem zeigte sich bei ähnlichen Problemen ein *Distanzeffekt* (Potts, 1974). Dieser Effekt zeigt, dass Konklusionen mit näher aneinander liegenden Termen (ArC) schwerer abgerufen werden als jene, bei denen die Terme

weiter voneinander entfernt sind (ArD). Auf der einen Seite würde die KRS Theorie vorhersagen, dass die Gruppen ebenfalls diesen Effekten unterliegen sollten. Auf der anderen Seite zeigen sich der Inferenz- und Distanzeffekt am deutlichsten in den Antwortzeiten, welche die Personen benötigen. Es ist denkbar, dass durch den Deliberationsprozess allgemein mehr Zeit notwendig wird als beim individuellen Denken und die Effekte deshalb nicht eindeutig nachweisbar sind. Dennoch ist in meinem Experiment 1 der Kontinuitätseffekt, der sich vor allem in der Lesezeit der Prämissen zeigt, ebenfalls bei Gruppen aufgetreten. So wäre es denkbar, dass auch die Gruppe dem Inferenz- und Distanzeffekt beim Denken unterliegt. Deshalb sollten diese Effekte in zukünftigen Experimenten ergänzend zum Kontinuitäts- und Unbestimmtheitseffekt untersucht werden.

6.5.2 Überzeugungsrevision in Gruppen

In meinen Experimenten zur Überzeugungsrevision habe ich keine direkte Gruppeninteraktion untersucht. Vielmehr habe ich ein künstliches Gruppensetting (Experiment 6) oder eine fiktive Diskussion (Experiment 7) erstellt und die Probanden haben diese ohne direkte Interaktionen auf den Computerbildschirm dargeboten bekommen. Doch wie würden reale Gruppen räumliche Überzeugungen revidieren? Ich nehme an, dass sich hier ebenfalls das sehr robuste LO-Prinzip als präferierte Revisionsstrategie zeigen wird. Es ist jedoch möglich, dass die Gruppe durch den Deliberationsprozess häufiger erkennt, dass die Revisionsalternativen gleichwertig sind. Deshalb nehme ich an, dass das LO-Prinzip zwar die vorherrschende Revisionsstrategie beim kollaborativen Revidieren sein sollte, sich aber seltener zeigt als beim Revidieren mentaler Modelle von Individuen. In zukünftigen Untersuchungen wäre außerdem interessant zu prüfen, ob das Quellenprinzip (Wolf et al., 2012) in echten Diskussionen manipuliert werden kann. So könnte man beispielweise Konföderierte zusammen mit Versuchspersonen diskutieren lassen. Einer der Konföderierten könnte sich so verhalten, dass er als vertrauenswürdiger empfunden wird und der andere sollte hingegen unzuverlässig wirken. Dann könnten die Konföderierten den Vpn Informationen (Prämissen) innerhalb eines Gesprächs mitteilen. Anschließend sollte man die Vpn mit einem inkonsistenten Fakt konfrontieren. Dabei ist anzunehmen, dass sich das Quellenprinzip durch das realistischere Setting im Vergleich zu meinem Experiment 7 deutlicher zeigt und die Vpn die Information von dem wenig vertrauenswürdigen Konföderierten verwerfen. Das experimentelle Paradigma mit Konföderierten wäre ebenfalls für eine weitere Überprüfung des Konformitätseffektes zum räumlichen Schlussfolgern interessant (siehe auch Kath, 2019). Bisher habe ich den Vpn lediglich die Informationen der (fiktiven) Gruppe auf dem

Computerbildschirm präsentiert. Es bleibt zu prüfen, ob eine Vpn ihre Revisionsstrategie auch der Gruppenmeinung anpasst, wenn sie in einer realen Gruppendiskussion ihre Revision verteidigen kann.

Besonders in realen Gruppendiskussionen beim Schlussfolgern sind Gruppenmitglieder fortwährend dazu gezwungen, ihre Überzeugungen zu revidieren. Deshalb sollte auch getestet werden, welche Faktoren innerhalb der Gruppe dazu beitragen, dass die Gruppenmitglieder bestimmte Überzeugungen der anderen Personen eher übernehmen. So wäre es möglich, dass ein Gruppenmitglied besonders kompetent für die zu lösende Denkaufgabe wirkt. Zusätzlich könnte man manipulieren, inwieweit die Demonstrierbarkeit eines Problems (Laughlin & Ellis, 1986) dazu führt, dass ein Gruppenmitglied die anderen Personen mehr beziehungsweise weniger erfolgreich überzeugen kann.

6.5.3 Gruppengröße und Reihenfolgeeffekte

Bisher habe ich das Denken in Gruppen ausschließlich an Dyaden getestet. Laughlin, Hatch, Silver und Boh (2006) zeigten beim Buchstaben-zu-Zahlen Problem, dass Gruppen aus zwei Personen nicht ausreichend waren, um die Leistung von Individuen zu übertreffen. Außerdem zeigten sie, dass Gruppen mit vier oder fünf Mitgliedern keine bessere Leistung als drei-Personen-Gruppen erzielen. Daraus schließen die Autoren, dass (1) eine Gruppe von drei Personen notwendig ist, um Individuen zu übertreffen und (2) eine Gruppe von drei Personen ausreichend ist, um die optimale Leistung zu erzielen. Diese Befunde konnte ich nicht stützen. In allen Experimenten zum Denken konnten die Dyaden insgesamt mehr valide Konklusionen ziehen als Individuen. Daraus lässt sich auf der einen Seite ableiten, dass für das Denken die Dyade (im Vergleich zum Problemlösen) bereits ausreichend ist, um die individuelle Leistung zu übertreffen. Auf der anderen Seite wirft es eine Frage auf: Reicht die Dyade aus, um die optimale Leistung beim Denken zu erzielen?

Die Befunde aus meinen Experimenten und die KRS Theorie lassen beide Interpretationen zu: So nehme ich einerseits an, dass mit jedem Gruppenmitglied mehr mentale Modelle in den Deliberationsprozess eingebracht werden. Dies sollte dazu führen, dass weniger Fehler entstehen und mehr valide Konklusionen gezogen werden können. Diese Annahme sollte auch für mehr als zwei Personen zutreffen. Andererseits vermute ich, dass der metakognitive Prozess des FOR nicht besonders durch die Gruppengröße beeinflusst wird. Ich nehme an, dass die Berücksichtigung des FOR dadurch gesteigert wird, dass man eine Gruppeninteraktion antizipiert. Dies sollte deshalb durch eine Gruppengröße von mehr als zwei Personen nicht erheblich gesteigert werden.

Bei Moshman und Geil (1998) konnten Gruppen bestehend aus fünf Personen die individuelle Leistung übertreffen. Die Autoren führen diese Überlegenheit auf den Diskussionsprozess zurück. Wenn jedoch, wie in der KRS Theorie angenommen, das FOR die Überlegenheit erklärt, dann sollte bereits eine Dyade ausreichend sein und sogar die Leistung von fünf-Personen-Gruppen erreichen können. Diese Annahmen müssen jedoch zunächst einer experimentellen Prüfung standhalten. Falls die Dyade noch nicht zur optimalen Leistung führen kann, bleibt zu klären, wie viele Personen einen Beitrag zum Denken leisten können. Wenn bereits beim Problemlösen eine Gruppe von drei Personen ausreichend ist (Laughlin et al., 2006), ist es abwegig, dass beim Denken in einer Gruppe deutlich mehr als drei Personen für die optimale Leistung notwendig wären.

Für zukünftige Experimente ist es außerdem sehr wichtig, das Denken von Individuen und Gruppen in abwechselnden Reihenfolgen zu testen. In Experiment 5 konnte ich zeigen, dass Individuen durch das vorherige Denken in der Gruppe profitieren konnten. Ich kann aber nicht eindeutig klären, wie stark dieser Einfluss ist und ob dieser auf metakognitive Prozesse wie das FOR zurückzuführen ist. Laughlin, Carey und Kerr (2008) testeten verschiedene Reihenfolgen von Individuen (I) und Gruppen (G) beim Problemlösen. Sie ließen die Vpn in Abfolgen in Form von „GGGG“, „GGGI“, „GGII“, „GIII“ oder „IIII“ verschiedene Probleme lösen. Sie konnten zeigen, dass eine einmalige Gruppeninteraktion ausreichte, um die Leistung auf das individuelle Problemlösen zu übertragen (GIII) und dass weitere Gruppeninteraktionen die individuelle Leistung nicht noch weiter steigern konnten (GGII). Dies wirft die Frage auf, ob der metakognitive Prozess des FOR ebenfalls durch einmalige Gruppeninteraktionen soweit verstärkt wird, dass er bei jedem darauffolgenden individuellen Denken ebenfalls stärker berücksichtigt wird. Es könnte aber auch sein, dass das FOR mit jedem weiteren individuellen Durchgang weniger Einfluss auf die Denkprozesse hat. Allgemein lässt das Testen der Probanden in unterschiedlichen Reihenfolgen wichtige weitere Rückschlüsse zum Denken in Gruppen zu. So könnte man eine Art Kombination aus der Methodik von Thompson et al. (2011) und Laughlin et al. (2008) vornehmen. Dabei sollten die Probanden zunächst eine schnelle, intuitive Konklusion angeben. Nach dieser Konklusion geben die Vpn an, wie sicher sie sich mit dieser Antwort sind (das Maß für FOR). Anschließend könnte man variieren, ob die mögliche Revision der ursprünglichen Konklusion alleine oder in einer Gruppe durchgeführt werden soll. Ich schlage vor, ein Experiment mit drei Bedingungen durchzuführen: „II, IG und GI“. Nehmen wir als Beispiel einen Durchgang, bei denen die Vpn bei einem indeterminierten Problem zunächst die Annotation nicht berücksichtigen konnten und deshalb ein niedriges FOR für ihre Konklusion verspürten: Die

Bedingung „II“ wäre analog zu den Experimenten aus Thompson et al. (2011). Nachdem die Vpn ihre ursprüngliche Konklusion und das FOR angegeben haben, sollten sie im zweiten Durchgang häufiger ihre ursprüngliche Konklusion revidieren. In der Bedingung „IG“ geben die Vpn auch zunächst alleine eine intuitive Konklusion. Darauf geben sie an, wie sicher sie sich mit dieser Konklusion sind. Nun könnte man in der darauffolgenden Gruppenphase messen, (1) wie häufig eine Person diese Unsicherheit in der Gruppe anspricht und (2) wie häufig die Konklusion der Gruppe von den intuitiven Konklusionen der Gruppenmitglieder abweicht. So könnte man prüfen, ob das subjektiv empfundene FOR mit einer häufigeren Diskussion der Konklusion in der Deliberationsphase assoziiert ist und wie häufig die Gruppenmitglieder ihre ursprünglichen Konklusionen revidieren. In der Bedingung „GI“ würde die Gruppe gemeinsam eine möglichst schnelle, intuitive Konklusion geben und ihre Gruppenmitglieder könnten einzeln ihr subjektiv empfundenes FOR angeben. Nun könnte man prüfen, ob die Vpn in der darauffolgenden individuellen Phase häufiger ihre Konklusion ändern oder an der ursprünglichen Gruppenmeinung festhalten. Ich erwarte, dass in diesem Fall (GI) die Gruppenmeinung zu einem hohen FOR führt und Konklusionen seltener in der zweiten individuellen Phase revidiert werden, als wenn beide Phasen individuell (II) durchgeführt werden. Natürlich wären in diesem Paradigma viele weitere Möglichkeiten (z. B. IGI oder GIII) denkbar. Jedoch sollten zunächst die grundsätzlichen Einflüsse der Gruppeninteraktion auf das FOR in einfachen Reihenfolgen getestet werden.

6.5.4 Soziale Entscheidungsschemata

Ich habe in meinem Experiment nicht nur die drei theoretischen Modelle der Entscheidungsschemata *truth wins*, *error wins* und *equiprobability* getestet, sondern auch ein eigenes Modell aus den beobachteten Daten abgeleitet. Das *collaboration wins* Modell zeigt, dass Gruppen beim Denken über das *truth wins* Modell hinaus die individuelle Leistung übertreffen können. Doch was bedeutet dies für zukünftige Forschungen? Zunächst muss das *collaboration wins* Modell in weiteren Experimenten getestet werden. Dabei sollte geprüft werden, ob es die Gruppenleistung beim Denken vorhersagen kann und ob es eine bessere Erklärung als das *truth wins* oder andere Modelle bieten kann. Es ist jedoch wichtig, die Modelle auch mit anderen methodischen Variationen zu testen. Eine der wichtigsten Fragen dabei lautet: Hält das *collaboration wins* Modell auch dann, wenn die Gruppe aus mehr als zwei Personen besteht? Auf theoretischer Ebene ist dies anzunehmen. Jedoch zeigte mein Experiment zur räumlichen Überzeugungsrevision und Konformität, dass die Majorität einen Einfluss auf das Denken der Einzelnen haben kann. Deshalb sollte in weiteren Experimenten

die Gruppengröße systematisch variiert werden und die von mir getesteten Modelle um weitere Modelle wie das *majority wins* (Davis, 1973) ergänzt werden. Auch andere Faktoren wie der Zeitdruck könnten zu interessanten Ergebnissen führen. So ist es denkbar, dass die Gruppe einem anderen Entscheidungsschema folgt (z. B. *equiprobability*), wenn ihnen nur begrenzte Zeit für den Deliberationsprozess zur Verfügung steht. Insgesamt können die sozialen Entscheidungsschemata uns wichtige Erkenntnisse über das Denken in Gruppen liefern. Dabei ist jedoch zu beachten, dass diese Modelle eine Idealvorstellung der Gruppenprozesse annehmen. So gehen sie davon aus, dass jedes Individuum mit gleicher Wahrscheinlichkeit die Lösung eines Problems finden kann. Außerdem treffen die Modelle keine Aussagen darüber, ob eine bestimmte Person mehr Bedeutung für die Gruppe aufweist. Deshalb betonen Kerr, Niedermeier und Kaplan (2000) dass nicht nur die Modelle mit sehr guter „Passung“ uns etwas über die Gruppenprozesse verraten können. Sie empfehlen, auch theoretisch unplausible Modelle zu testen. Wenn zum Beispiel das *collaboration wins* die beobachtete Gruppenleistung nicht erheblich besser als ein unplausibles Modell vorhersagen kann, dann sollte das *collaboration wins* Modell keine starke Aussagekraft über das Denken in Gruppen haben. In diesem Fall müssen gegebenenfalls dynamischere Modelle, die Einflussfaktoren wie Dominanz (Cummins, 2005) oder Expertise (Stevenson & Over, 2001) einzelner Gruppenmitglieder berücksichtigen können, entwickelt werden.

6.5.5 Synergieeffekte beim kollaborativen Denken

Können bestimmte Umstände oder Interventionen die Gruppenleistung beim Schlussfolgern beeinflussen? Nach Larson (2010) weisen Gruppen, die in hohem Maße effektiv Probleme lösen, eine hohe *Synergie* auf. Doch wie kann man Synergie beim kollaborativen Schlussfolgern erzeugen? Dillenbourg und Traum (2006) zeigten, dass man durch mediale Unterstützung die Gruppenleistung beeinflussen kann. Sie ließen Gruppen über ein Computersystem mit einem Sprachchat einen Kriminalfall lösen, bei dem verschiedene Verdächtige und Zeugen sich in unterschiedlichen Räumen einer virtuellen Umgebung aufhielten. Damit man den Mörder entlarven konnte, mussten die Gruppenmitglieder ihre Repräsentationen über die Verdächtigen in den verschiedenen Räumen untereinander austauschen. Dazu hatten die Vpn aber nicht nur die Sprache als Medium zur Verfügung. Zusätzlich durften sie ein virtuelles Whiteboard verwenden. Die Autoren nahmen an, dass die Vpn das Whiteboard lediglich dazu verwenden, Unklarheiten bei der verbalen Vermittlung einer Repräsentation aufzulösen. Das Experiment zeigte jedoch, dass das Whiteboard der

Sprache überlegen war. Die Vpn verwendeten das Whiteboard, um ihre mentalen Modelle zu konkretisieren und die Sprache diente lediglich dazu, Unklarheiten aufzulösen.

In fast allen meiner Experimente durften die Vpn ihre Modelle nur verbal miteinander austauschen. In Experiment 3 haben die Vpn Legoblocke verwendet, um die mentalen Modelle miteinander zu kommunizieren. In diesem Experiment wurden die Individuen mit steigender Schwierigkeit der Probleme immer schlechter. Die Leistung der Gruppe hingegen verschlechterte sich nicht. Das heißt, dass die Gruppe möglicherweise durch das Mitteilen der Repräsentationen durch reale Objekte eine bessere Denkleistung erzielen konnte. Dabei ist jedoch anzumerken, dass das Ablesen einer Konklusion nur mental stattfinden kann. Wenn die Vpn reale Objekte auf den Tisch legen, misst man lediglich, wie gut die Vpn (gemeinsam) ein (geteiltes) Modell konstruieren können. Dieses methodische Problem könnte durch ein virtuelles Whiteboard wie bei Dillenbourg und Traum (2006) gelöst werden, indem das Whiteboard vor dem Abfragen einer möglichen Konklusion gelöscht wird. Es sollte also in weiteren Experimenten getestet werden, welches Medium für das Denken der Gruppe hilfreich sein könnte oder ob es Medien gibt, die sogar für das Denken in Gruppen hinderlich sein können.

Eine weitere Möglichkeit das Denken in Gruppen zu erleichtern, ist das Spezifizieren von verschiedenen Rollen für die Gruppenmitglieder. Schweiger, Sandberg und Ragan (1986) zeigten bereits beim Entscheiden, dass das Einführen eines *Teufelsadvokaten* (Herbert & Estes, 1977) hilfreich sein kann. Der Teufelsadvokat ist eine besondere Rolle, die alle Aussagen der anderen Gruppenmitglieder in Frage stellt. Dabei fordert der Advokat die anderen Gruppenmitglieder auf, ihre Argumente so genau wie möglich zu erklären und zu verteidigen. Diese Rolle könnte auch für das kollaborative räumliche Schlussfolgern bedeutsam sein. Wenn eine Person – als Teufelsadvokat – die Gruppenmitglieder auffordert, ihre mentalen Modelle und Konklusionen so konkret wie möglich zu erläutern, dann könnte dies dazu führen, dass die Gruppenmitglieder häufiger mehrere Modelle berücksichtigen und die Gruppenleistung gesteigert wird.

In weiteren Untersuchungen sollte außerdem getestet werden, inwieweit verschiedene Kommunikationsstrategien einen Synergieeffekt fördern können. Bei der Nominalen-Gruppentechnik (*nominal group technique*; Van de Ven & Delbecq, 1972, 1974) müssen die Gruppenmitglieder in einem strukturierten Prozess zunächst individuell ihre Ideen notieren und dann (ohne Diskussion) diese der Gruppe vorstellen. Danach werden alle Vorschläge diskutiert und eine Gruppenantwort abgeleitet. Diese Technik ist eine effektive Methode, um Gruppendiskussionen zu operationalisieren und zu verbessern (McMillan, King, & Tully,

2016). Deshalb nehme ich an, dass sich diese Technik auch für das kollaborative Schlussfolgern als hilfreich erweisen sollte. So wäre sichergestellt, dass jedes Gruppenmitglied die Möglichkeit hat, ihr eigenes mentales Modell zu präsentieren und gegebenenfalls die Annotation über weitere Interpretationen der Prämissen zu äußern. Diese Annahmen werden dann erst am Ende diskutiert und in ein geteiltes mentales Modell überführt. Diese Technik könnte also sicherstellen, dass wirklich alle individuellen mentalen Modelle berücksichtigt werden und nicht einige der Gruppenmitglieder vernachlässigt werden.

6.6 Fazit

Ich konnte mit meiner Arbeit zeigen, dass eine individualistische Sicht nicht ausreichend ist, um die gesamten Kapazitäten des menschlichen Denkens zu erklären.

Gruppen können gemeinsam eine geteilte Repräsentation von räumlich-relationalen Problemen konstruieren und daraus Informationen ablesen. Dabei sind sie in der Lage, über das Denken von Individuen hinaus folgerichtige Schlussfolgerungen zu ziehen. Das *collaboration wins* Modell zeigt, dass die Gruppenleistung über das einfache Austauschen der Repräsentationen hinaus das Denken der Individuen übersteigen kann. Außerdem habe ich gezeigt, dass sogar das Denken mit räumlichen Inhalten durch gruppenkonformes Verhalten beeinflusst wird. Wie aber in jeder wissenschaftlichen Arbeit entstanden durch jede Frage, die ich mit meinen Experimenten beantworten konnte, mindestens genauso viele neue Fragen. Inwieweit kann man das kollaborative Denken verbessern und welche Rolle spielt die Gruppengröße beim Denken? Unterliegen die Gruppen auch anderen bekannten Effekten und wann stößt die Gruppe beim Denken an ihre Grenzen? Die Beantwortung dieser Fragen hat nicht nur große Relevanz für die Denk- und Sozialpsychologie, sondern ist auch für das alltägliche Denken wichtig. Eine Gesellschaft, die sich in hohem Maße auf das Denken von Gruppen verlässt, sollte genau wissen, wie Gruppen denken und wie diese das menschliche Denken schärfen können.

Einerseits hatte Dr. Watson also Recht über seine Bedeutung für das Denken von Sherlock Holmes. Menschen denken nicht isoliert von anderen Personen und benötigen die Interaktion, um ihr wahres Potential zu erreichen. Andererseits sind die Aussagen von Dr. Watson ein wenig zu simpel. Das menschliche Denken ist ein komplexer Prozess und selbst auf individueller Ebene können wir noch nicht alle Prozesse erklären. Deshalb sollten wir – genau wie Dr. Watson – alles mit Einwüfen bedenken, unser Denken gegenseitig schärfen und dieses noch sehr junge Forschungsgebiet kollaborativ erschließen.

Kapitel 7: Literatur

- Alchourrón, C. E., Gärdenfors, P., & Makinson, D. (1985). On the logic of theory change: Partial meet contraction and revision functions. *The Journal of Symbolic Logic*, 50(2), 510-530.
- Anderson, J. R. (2009). *Cognitive psychology and its implications*. New York: Worth Publishers.
- Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgments. In H. Guetzkow (Ed.), *Groups, leadership and men: Research in human relations* (pp. 177-190). Oxford: Carnegie Press.
- Asch, S. E. (1956). Studies of independence and conformity. A minority of one against a unanimous majority. *Psychological Monographs*, 70(9), 1-70.
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *Psychology of Learning and Motivation*, 2, 89-195.
- Baddeley, A. (1992). Working memory. *Science*, 255(5044), 556-559.
- Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (pp. 47-89). Academic Press
- Ball, L. J., Phillips, P., Wade, C. N., & Quayle, J. D. (2006). Effects of belief and logic on syllogistic reasoning: Eye-movement evidence for selective processing models. *Experimental Psychology*, 53(1), 77-86.
- Barrouillet, P., Gauffroy, C., & Lecas, J. F. (2008). Mental models and the suppositional account of conditionals. *Psychological Review*, 115(3), 760-771.
- Begg, I., & Harris, G. (1982). On the interpretation of syllogisms. *Journal of Verbal Learning and Verbal Behavior*, 21(5), 595-620.
- Bonnefon, J. F., Girotto, V., & Legrenzi, P. (2012). The psychology of reasoning about preferences and unsequential decisions. *Synthese*, 185(1), 27-41.
- Bortz, J., & Schuster, C. (2010). *Statistik für Human-und Sozialwissenschaftler*. Berlin: Springer.
- Bowman, J. M., & Wittenbaum, G. M. (2012). Time pressure affects process and performance in hidden-profile groups. *Small Group Research*, 43(3), 295-314.
- Brain und O'Brien (1991). A theory of *if*: A lexical entry, reasoning program, and pragmatic principles. *Psychological Review*, 98(2), 182-203.
- Brauner, E., & Scholl, W. (2000). Editorial: The information processing approach as a perspective for groups research. *Group Processes & Intergroup Relations*, 3(2), 115-122.
- Bucher, L., Krumnack, A., Nejasnic, J., & Knauff, M. (2011). Cognitive processes underlying spatial belief revision. In *Proceedings of the Annual Conference of the Cognitive Science Society* (Vol. 35, pp. 3477-3483). Austin, TX: Cognitive Science Society.

- Bucher, L., & Nejasmic, J. (2012, August). Relocating multiple objects during spatial belief revision. In C. Stachniss, K. Schill, & D. Uttal (Eds.), *International Conference on Spatial Cognition* (pp. 478-491). Berlin, Heidelberg: Springer.
- Bucher, L., Nejasmic, J., Bertleff, S., & Knauff, M. (2013). Plausibility and visualizability in relational belief revision. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 35, pp. 1946-1951). Austin, TX: Cognitive Science Society.
- Bucher, L., Röser, F., Nejasmic, J., & Hamburger, K. (2014). Belief revision and way-finding. *Cognitive Processing*, 15(1), 99-106.
- Bucher, L., & Thorn, P. (2014). Minimality criteria in spatial belief revision. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 36, pp. 1952-1957). Austin, TX: Cognitive Science Society.
- Bullock, M., & Sodian, B. (2003). Entwicklung des wissenschaftlichen Denkens. In W. Schneider & M. Knopf (Eds.), *Entwicklung, Lehren und Lernen* (pp. 75-92). Göttingen: Hogrefe.
- Byrne, R. M., & Johnson-Laird, P. N. (1989). Spatial reasoning. *Journal of Memory and Language*, 28(5), 564-575.
- Carreiras, M., & Santamaría, C. (1997). Reasoning about relations: Spatial and nonspatial problems. *Thinking & Reasoning*, 3(3), 191-208.
- Chater, N., & Oaksford, M. (2012). Normative systems: Logic, probability, and rational choice. In K. Holyoak & R. G. Morrison (Eds.) *The Oxford handbook of thinking and reasoning* (pp. 11-21). Oxford: Oxford University Press.
- Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the wason selection task. *Cognition*, 31(3), 187-276.
- Cosmides, L., & Tooby, J. (2000). The cognitive neuroscience of social reasoning. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp. 1259-1279). Cambridge: MIT Press.
- Cosmides, L., Tooby, J., Fiddick, L., & Bryant, G. A. (2005). Detecting cheaters. *Trends in Cognitive Sciences*, 9(11), 505-506.
- Craik, K. (1943). *The nature of explanation*. Cambridge: Cambridge University Press.
- Cummins, D. (2005). Dominance, status, and social hierarchies. In D. M. Buss (Ed.), *The handbook of evolutionary psychology* (pp. 676-697). Hoboken, NJ: Wiley
- Davis (1973). Group decision and social interaction: A theory of social decision schemes. *Psychological Review*, 80(2), 97-125.
- Davis (1996). Group decision making and quantitative judgments: A consensus model. In E. Witte & H. Davis (Eds.), *Understanding group behavior, Vol. 1: Consensual action by small groups* (pp. 35-39). Hillsdale: Lawrence Erlbaum.
- De Neys, W. (2014). Conflict detection, dual processes, and logical intuitions: Some clarifications. *Thinking & Reasoning*, 20(2), 169-187.

- De Neys, W., Moyens, E., & Vansteenwegen, D. (2010). Feeling we're biased: Autonomic arousal and reasoning conflict. *Cognitive, Affective, & Behavioral Neuroscience*, *10*(2), 208-216.
- De Neys, W., Schaeken, W., & D'Ydewalle, G. (2003). Inference suppression and semantic memory retrieval: Every counterexample counts. *Memory & Cognition*, *31*(4), 581-595.
- Denis, M., & Cocude, M. (1989). Scanning visual images generated from verbal descriptions. *European Journal of Cognitive Psychology*, *1*(4), 293-307.
- De Soto, C. B., London, M., & Handel, S. (1965). Social reasoning and spatial paralogic. *Journal of Personality and Social Psychology*, *2*(4), 513-521.
- Dillenbourg, P. & Betrancourt, M. (2006). Collaboration load: Handling complexity in learning environments: Theory and research. In J. Elen & E. Clark (Eds.), *Advances in learning and instruction series* (pp. 142-163). Oxford: Elsevier.
- Dillenbourg, P., & Traum, D. (2006). Sharing solutions: Persistence and grounding in multimodal collaborative problem solving. *The Journal of the Learning Sciences*, *15*(1), 121-151.
- Doyle, A. C. (1993). The adventures of the creeping man. In A. C. Doyle (Ed.), *The case-book of Sherlock Holmes* (pp. 31-44). Hertfordshire: Wordsworth Editions.
- Dunbar, R. I. (1998). The social brain hypothesis. *Evolutionary anthropology. Issues, News, and Reviews*, *6*(5), 178-190.
- Dunbar, K., & Fugelsang, J. (2006). Problem solving and reasoning. In E. E. Smith & S. M. Kosslyn (Eds.), *Cognitive psychology: Mind and brain*. New York: Prentice Hall.
- Ehrlich, K., & Johnson-Laird, P. N. (1982). Spatial descriptions and referential continuity. *Journal of Verbal Learning and Verbal Behavior*, *21*(3), 296-306.
- Elio, R., & Pelletier, F. J. (1997). Belief change as propositional update. *Cognitive Science*, *21*(4), 419-460.
- Evans, J. S. B. T. (2017). A brief history of the wason selection task. In N. Galbraith, E. Lucas, & D. E. Over (Eds.), *The thinking mind: A festschrift for Ken Manktelow* (pp. 1-14). New York, NY, US: Routledge/Taylor & Francis Group.
- Evans, J. S. B. T., Newstead, S. E., & Byrne, R. M. J. (1993). *Human reasoning: The psychology of deduction*. Hove, UK: Lawrence Erlbaum Associates.
- Evans, J. S. B. T., Over, D. E., & Handley, S. J. (2005). Suppositions, extensionality, and conditionals: A critique of the mental model theory of Johnson-Laird and Byrne (2002). *Psychological Review*, *112*(4), 1040-1052.
- Fermé, E., & Hansson, S. O. (2011). AGM 25 years. *Journal of Philosophical Logic*, *40*(2), 295-331.
- Festinger (1957). *A theory of cognitive dissonance*. Stanford: Stanford University Press
- Fisher, M., Knobe, J., Strickland, B., & Keil, F. C. (2017). The influence of social interaction on intuitions of objectivity and subjectivity. *Cognitive Science*, *41*(4), 1119-1134.

- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34*(10), 906-911.
- Gawronski, B., & Strack, F. (2012). *Cognitive consistency: A fundamental principle in social cognition*. New York: Guilford Press.
- Gazzo Castañeda, L. E., & Knauff, M. (2013). Individual differences, imagery and the visual impedance effect. In *Proceedings of the Annual Conference of the Cognitive Science Society* (Vol 35., pp. 2374-2379). Austin, TX: Cognitive Science Society.
- Gentzen, G. (1935). Untersuchungen über das logische Schließen. I. *Mathematische Zeitschrift*, *39*(1), 176-210.
- Gigerenzer, G., & Hug, K. (1992). Domain-specific reasoning: Social contracts, cheating, and perspective change. *Cognition*, *43*(2), 127-171.
- Gigone, D. (2010). Common knowledge effect. In J. M. Levine & M. A. Hogg (Eds.), *Encyclopedia of group processes & intergroup relations* (pp. 124-125). Thousand Oaks, CA: SAGE Publications, Inc.
- Goodwin, G. P., & Johnson-Laird, P. N. (2005). Reasoning about relations. *Psychological Review*, *112*(2), 468-493.
- Greitemeyer, T., & Schulz-Hardt, S. (2003). Preference-consistent evaluation of information in the hidden profile paradigm: Beyond group-level explanations for the dominance of shared information in group decisions. *Journal of Personality and Social Psychology*, *84*(2), 322-339.
- Grice, H. P. (1975). Logic and conversation. In M. Ezcurdia & R. J. Stainton (Eds.), *The semantics-pragmatics boundary in philosophy* (pp. 47-59). London: Broadview Press.
- Haas G. (2019) Formale Repräsentation von Überzeugungen und Theorien der Überzeugungsrevision. In M. Grajner & G. Melchior (Eds.), *Handbuch Erkenntnistheorie*. Stuttgart: Metzler.
- Hagert, G. (1985). What's in a mental model? On conceptual models in reasoning with spatial descriptions. In *Proceedings of the 9th International Joint Conference Artificial Intelligence (IJCAI)* (pp. 274-277). San Francisco, CA.
- Hagert, G., & Hansson, Å. (1983). *Logic modelling of cognitive reasoning*. Uppsala Programming Methodology and Artificial Intelligence Laboratory, Department of Computing Science, Uppsala University.
- Hahn, U., & Oaksford, M. (2007). The rationality of informal argumentation: A Bayesian approach to reasoning fallacies. *Psychological Review*, *114*(3), 704-732.
- Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences*, *21*(6), 803-831.
- Halford, G. S., Wilson, W. H., & Phillips, S. (2010). Relational knowledge: The foundation of higher cognition. *Trends in Cognitive Sciences*, *14*(11), 497-505.

- Hamburger, K., Ragni, M., Karimpur, H., Franzmeier, I., Wedell, F., & Knauff, M. (2018). TMS applied to V1 can facilitate reasoning. *Experimental Brain Research*, 236(8), 2277-2286.
- Harman, G. (1986). *Change in view: Principles of reasoning*. Cambridge, MA: The MIT Press.
- Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology*, 56(4), 208-216.
- Hasson, U., & Johnson-Laird, P. N. (2003). Why believability cannot explain belief revision. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 25, pp. 516-521). Austin, TX: Cognitive Science Society.
- Hastie, R., & Kameda, T. (2005). The robust beauty of majority rules in group decisions. *Psychological Review*, 112(2), 494-508.
- Hayward, W. G., & Tarr, M. J. (1995). Spatial language and spatial representation. *Cognition*, 55(1), 39-84.
- Herbert, T. T., & Estes, R. W. (1977). Improving executive decisions by formalizing dissent: The corporate devil's advocate. *Academy of Management Review*, 2(4), 662-667.
- Hinsz, V. B. (2004). Metacognition and mental models in groups: An illustration with metamemory of group recognition memory. In E. Salas & S. M. Fiore (Eds.), *Team cognition: Understanding the factors that drive process and performance* (p. 33-58). American Psychological Association.
- Hinsz, V. B., Tindale, R. S., & Vollrath, D. A. (1997). The emerging conceptualization of groups as information processors. *Psychological Bulletin*, 121(1), 43-64.
- Hunter, I. M. L. (1957). The solving of three-term series problems. *British Journal of Psychology*, 48(4), 286-298.
- IBM Corp. (2016). *IBM SPSS Statistics for Windows*, Version 24.0. Armonk, NY: IBM Corp.
- Ingham, A. G., Levinger, G., Graves, J., & Peckham, V. (1974). The ringelmann effect: Studies of group size and group performance. *Journal of Experimental Social Psychology*, 10(4), 371-384.
- Jahn, G., Knauff, M., & Johnson-Laird, P. N. (2007). Preferred mental models in reasoning about spatial relations. *Memory & Cognition*, 35(8), 2075-2087.
- James, W. (1890). Series of even difference and mediate comparison (Kapitel 28). In *Delphi complete works of William James* [Kindle Version, 2018]. Abgerufen von <http://www.amazon.com/>
- Janis, I. L. (1972). *Victims of groupthink: A psychological study of foreign-policy decisions and fiascoes*. Boston: Houghton Mifflin
- Janis, I. L. (1982). *Groupthink: Psychological studies of policy decisions and fiascoes*. Boston: Houghton Mifflin.
- Janis, I. L. (2008). Groupthink. *IEEE Engineering Management Review*, 36(1), 36.

- Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press.
- Johnson-Laird, P. N., & Bara, B. G. (1984). Syllogistic inference. *Cognition*, 16(1), 1-61.
- Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hove, UK: Lawrence Erlbaum Associates.
- Johnson-Laird, P. N., Girotto, V., & Legrenzi, P. (2004). Reasoning from inconsistency to consistency. *Psychological Review*, 111(3), 640-661.
- Johnson-Laird, P. N., & Wason, P. C. (1970). A theoretical analysis of insight into a reasoning task. *Cognitive Psychology*, 1(2), 134-148.
- Kameda, T., Tindale, R. S., & Davis, J. H. (2003). Cognitions, preferences, and social sharedness: Past, present, and future directions in group decision making. In S. L. Schneider & J. Shanteau (Eds.), *Emerging perspectives in judgment and decision research* (pp. 458-485). Cambridge: Cambridge University Press.
- Karacora, B., Dehghani, M., Kramer-Mertens, N., & Gratch, J. (2012). The influence of virtual agents' gender and rapport on enhancing math performance. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 34, pp. 563-568). Austin, TX: Cognitive Science Society.
- Kath, B. S. (2019). *Der Einfluss der Gruppengröße auf die Revision räumlicher Überzeugungen – Eine experimentelle Studie* (unpublished master thesis). Justus-Liebig-Universität Gießen, Gießen.
- Katsuno, H. & Mendelzon, A. (1991). Propositional knowledge base revision and minimal change. *Artificial Intelligence*, 52(3), 263-294.
- Kaup, B. & Dudschig, C. (2017). Sätze und Texte verstehen und produzieren. In J. Müsseler & M. Rieger (Eds.), *Allgemeine Psychologie*. Springer, Berlin, Heidelberg.
- Kerr, N. L., & Tindale, R. S. (2004). Group performance and decision making. *Annual Review of Psychology*, 55(1), 623-655.
- Kerr, N. L., Niedermeier, K. E., & Kaplan, M. F. (2000). On the virtues of assuming minimal differences in information processing in individuals and groups. *Group Processes & Intergroup Relations*, 3(2), 203-217.
- Kießner, A. K., & Ragni, M. (2017, July). Processing Spatial Relations: A meta-analysis. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 29, pp. 16-19). Austin, TX: Cognitive Science Society.
- Klauer, K. C., & Singmann, H. (2013). Does logic feel good? Testing for intuitive detection of logicity in syllogistic reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(4), 1265-1273.
- Knauff, M. (2006). Deduktion und logisches Denken. In J. Funke (Ed.), *Enzyklopädie der Psychologie, Themenbereich C, Bd. 8, Denken und Problemlösen* (pp. 167-264). Göttingen, Germany: Hogrefe.

- Knauff, M. (2013). *Space to reason: A spatial theory of human thought*. Cambridge, MA: The MIT Press.
- Knauff, M., Bucher, L., Krumnack, A., & Nejasmic, J. (2013). Spatial belief revision. *Journal of Cognitive Psychology*, 25(2), 147-156.
- Knauff, M., & Johnson-Laird, P. N. (2002). Visual imagery can impede reasoning. *Memory & Cognition*, 30(3), 363-371.
- Knauff M., & Knoblich G. (2017) Logisches Denken. In J. Müsseler & M. Rieger (Eds.), *Allgemeine Psychologie*. Springer, Berlin, Heidelberg.
- Knauff, M., Rauh, R., & Schlieder, C. (1995). Preferred mental models in qualitative spatial reasoning: A cognitive assessment of Allen's calculus. In *Proceedings of the Annual Conference of the Cognitive Science Society* (Vol. 17, pp. 200-205). Mahwah, NJ: Lawrence Erlbaum Associates.
- Knauff, M., Rauh, R., Schlieder, C., & Strube, G. (1998a). Continuity effect and figural bias in spatial relational inference. In *Proceedings of the Annual Conference of the Cognitive Science Society* (Vol. 20, pp. 573-578). Mahwah, NJ: Lawrence Erlbaum Associates.
- Knauff, M., Rauh, R., Schlieder, C., & Strube, G. (1998b). Mental models in spatial reasoning. In C. Freksa, C. Habel, & K. F. Wender (Eds.), *Spatial cognition* (pp. 267-291). Berlin, Germany: Springer.
- Kolfschoten, G. L., & Brazier, F. M. (2013). Cognitive load in collaboration: Convergence. *Group Decision and Negotiation*, 22(5), 975-996.
- Koriat, A. (2007). Metacognition and consciousness. In P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The Cambridge handbook of consciousness* (pp. 289-326). Cambridge, NY: Cambridge University Press.
- Krumnack, A., Bucher, L., Nejasmic, J., & Knauff, M. (2011). Efficiency and minimal change in spatial belief revision. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 33, pp. 2270-2275). Austin, TX: Cognitive Science Society.
- Landau, B., & Jackendoff, R. (1993). "What" and "where" in spatial language and spatial cognition. *Behavioral and Brain Sciences*, 16(2), 217-238.
- Larson (2010). *In search of synergy: In small group performance*. New York: Psychology Press.
- Latané, B., Williams, K., & Harkins, S. (1979). Many hands make light the work: The causes and consequences of social loafing. *Journal of Personality and Social Psychology*, 37(6), 822-832.
- Laughlin, P. R. (1980). Social combination processes of cooperative problem-solving groups on verbal intellectual tasks. In M. Fishbein (Ed.), *Progress in Social Psychology* (pp. 127-155). Hillsdale, NJ: Erlbaum.
- Laughlin, P. R., Bonner, B. L., & Miner, A. G. (2002). Groups perform better than the best individuals on letters-to-numbers problems. *Organizational Behavior and Human Decision Processes*, 88(2), 605-620.

- Laughlin, P. R., Carey, H. R., & Kerr, N. L. (2008). Group-to-individual problem-solving transfer. *Group Processes & Intergroup Relations*, 11(3), 319-330.
- Laughlin, P. R., & Ellis, A. L. (1986). Demonstrability and social combination processes on mathematical intellectual tasks. *Journal of Experimental Social Psychology*, 22(3), 177-189.
- Laughlin, P. R., Hatch, E. C., Silver, J. S., & Boh, L. (2006). Groups perform better than the best individuals on letters-to-numbers problems: Effects of group size. *Journal of Personality and Social Psychology*, 90(4), 644-651.
- Laughlin, P. R., Zander, M. L., Kniewel, E. M., & Tan, T. K. (2003). Groups perform better than the best individuals on letters-to-numbers problems: Informative equations and effective strategies. *Journal of Personality and Social Psychology*, 85(4), 684-694.
- Leiner, D. J. (2019). SoSci Survey (Version 3.1.06) [Computer software]. Available at <https://www.soscisurvey.de>
- List, C., & Pettit, P. (2006). Group agency and supervenience. *The Southern Journal of Philosophy*, 44(1), 85-105.
- List, C., & Pettit, P. (2011). *Group agency: The possibility, design, and status of corporate agents*. Oxford: Oxford University Press.
- Lorge, I., & Solomon, H. (1955). Two models of group behavior in the solution of eureka-type problems. *Psychometrika*, 20(2), 139-148.
- Luchins, A. S. & Luchins, E. H. (1959). *Rigidity of behavior*. Eugene: University of Oregon.
- Maciejovsky, B., & Budescu, D. V. (2007). Collective induction without cooperation? Learning and knowledge transfer in cooperative groups and competitive auctions. *Journal of Personality and Social Psychology*, 92(5), 854-870.
- Makinson, D., & Gärdenfors, P. (1988). Revisions of knowledge systems using epistemic entrenchment. In *Proceedings of the Second Conference on Theoretical Aspects of Reasoning about Knowledge (TARK88)* (pp. 83-95). Morgan Kaufmann Publishers Inc..
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York: Freeman and Company
- Mathias, A. P., Vogel, P., & Knauff, M. (2019). Different cognitive styles can affect performance in laparoscopic surgery skill training. *Surgical Endoscopy*, 1-8.
- Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, 44(2), 314-324.
- Matz, D. C., & Wood, W. (2005). Cognitive dissonance in groups: The consequences of disagreement. *Journal of Personality and Social Psychology*, 88(1), 22-37.
- McCroskey, J. C., & Teven, J. J. (1999). Goodwill: A reexamination of the construct and its measurement. *Communications Monographs*, 66(1), 90-103.
- McMillan, S. S., King, M., & Tully, M. P. (2016). How to use the nominal group and delphi techniques. *International Journal of Clinical Pharmacy*, 38(3), 655-662.

- Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, *34*(2), 57-74.
- Mercier, H., & Sperber, D. (2017). *The enigma of reason*. Cambridge: Harvard University Press.
- Microsoft Corporation. (2016). *Microsoft Excel*. Retrieved from <https://office.microsoft.com/excel>
- Moede, W. (1927). Die Richtlinien der Leistungspsychologie. *Industrielle Psychotechnik*, *4*, 193-209.
- Moorhead, G., Ference, R., & Neck, C. P. (1991). Group decision fiascoes continue: Space shuttle Challenger and a revised groupthink framework. *Human Relations*, *44*(6), 539-550.
- Moshman, D. & Geil, M. (1998). Collaborative reasoning: Evidence for collective rationality. *Thinking & Reasoning*, *4*(3), 231-248.
- Nejasmic, J., Bucher, L., & Knauff, M. (2015). The construction of spatial mental models – A new view on the continuity effect. *The Quarterly Journal of Experimental Psychology*, *68*(9), 1794-1812.
- Nejasmic, J., Bucher, L., Thorn, P. D., & Knauff, M. (2014). Construction and revision of spatial mental models under high task demand. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 36, pp. 1066-1071). Austin, TX: Cognitive Science Society.
- Neubaum, G., Rösner, L., Ganster, T., Hambach, K., & Krämer, N. C. (2018). United in the name of justice: How conformity processes in social media may influence online vigilantism. *Psychology of Popular Media Culture*, *7*(2), 185-199.
- Nir, A. (2004). *A behavioral model of consumption patterns: The effects of cognitive dissonance and conformity*. CentER Discussion Paper (Vol. 2004-2048). Tilburg: Microeconomics
- Oakhill J., & Garnham, A. (1985) Referential continuity, transitivity, and the retention of relational descriptions, *Language and Cognitive Processes*, *1*(2), 149-162.
- Oaksford, M. (2015). Imaging deductive reasoning and the new paradigm. *Frontiers in Human Neuroscience*, *9*(101), 1-14.
- Oaksford, M., & Chater, N. (1995). Theories of reasoning and the computational explanation of everyday inference. *Thinking & Reasoning*, *1*(2), 121-152.
- Oaksford, M., & Chater, N. (2019). New paradigms in the psychology of reasoning. *Annual Review of Psychology*, *71*, 305-330.
- Öllinger, M. (2017). Problemlösen. In J. Müsseler & M. Rieger (Eds.), *Allgemeine Psychologie* (pp. 587-618). Springer, Berlin, Heidelberg.
- Over, D. E. (2009). New paradigm psychology of reasoning. *Thinking & Reasoning*, *15*(4), 431-438.

- Perham, N., & Oaksford, M. (2005). Deontic reasoning with emotional content: Evolutionary psychology or decision theory? *Cognitive Science*, 29(5), 681-718.
- Piaget, J. (1965). *Logic and psychology*. Manchester: Manchester University Press.
- Potts, G. R. (1972). Information processing strategies used in the encoding of linear orderings. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 727-740.
- Potts, G. R. (1974). Storing and retrieving information about ordered relationships. *Journal of Experimental Psychology*, 103(3), 431-439.
- Ragni, M., Fangmeier, T., Webber, L., & Knauff, M. (2006). Complexity in spatial reasoning. In *Proceedings of the Annual Cognitive Science Conference* (Vol. 28, pp. 2017-2022). Austin, TX: Cognitive Science Society.
- Ragni, M., & Knauff, M. (2013). A theory and a computational model of spatial reasoning with preferred mental models. *Psychological Review*, 120(3), 561-588.
- Ragni, M & Knauff, M. (2013). Preferred inferences in reasoning with spatial mental models (PRISM). Abgerufen von <http://spatialmentalmodels.appspot.com/>.
- Ragni, M., Knauff, M., & Nebel, B. (2005). A computational model for spatial reasoning with mental models. In *Proceedings of the Annual Cognitive Science Conference* (Vol. 27, pp. 1797-1802). Mahwah, NJ: Lawrence Erlbaum Associates.
- Ragni, M. & Kola, I. & Johnson-Laird, P. (2017). The Wason Selection Task: A meta-analysis. In *Proceedings of the Annual Cognitive Science Conference* (Vol. 39, pp. 980 - 985). Austin, TX: Cognitive Science Society.
- Revlis, R., Lipkin, S. G., & Hayes, J. R. (1971). The importance of universal quantifiers in a hypothetical reasoning task. *Journal of Verbal Learning and Verbal Behavior*, 10(1), 86-91.
- Rauh, R. (2000). Strategies of constructing preferred mental models in spatial relational inference. In W. Schaeken, G. De Vooght, A. Vandierendonck & G. D'Ydewalle (Eds.), *Deductive reasoning and strategies* (pp. 177-190). Mahwah, NJ: Lawrence Erlbaum.
- Rauh, R., Hagen, C., Knauff, M., Kuß, T., Schlieder, C., & Strube, G. (2005). From preferred to alternative mental models in spatial reasoning. *Spatial Cognition & Computation*, 5(2-3), 239-269.
- Raven, J. C. (1998). *Raven's progressive matrices and vocabulary scales*. Oxford: Psychologists Press.
- Reason, J. (1987). The chernobyl errors. *Bulletin of the British Psychological Society*, 40(206), 1-20.
- Revlis, R., Lipkin, S. G., & Hayes, J. R. (1971). The importance of universal quantifiers in a hypothetical reasoning task. *Journal of Verbal Learning and Verbal Behavior*, 10(1), 86-91.
- Rips, L. J. (1994). *The psychology of proof: Deductive reasoning in human thinking*. Cambridge, MA: The MIT Press.
- Roberts, M. J. (2000). Strategies in relational inference. *Thinking & Reasoning*, 6(1), 1-26.

- Rott, H. (1992). Preferential belief change using generalized epistemic entrenchment. *Journal of Logic, Language and Information*, 1(1), 45-78.
- Rott, H. (2001). *Change, choice and inference: A study of belief revision and nonmonotonic reasoning* (No. 42). Oxford University Press.
- Schneider, W. (Ed.). (2008). *Entwicklung von der Kindheit bis zum Erwachsenenalter: Befunde der Münchner Längsschnittstudie LOGIK*. Beltz, PVU.
- Scholten, L., Van Knippenberg, D., Nijstad, B. A., & De Dreu, C. K. (2007). Motivated information processing and group decision-making: Effects of process accountability on information processing and decision quality. *Journal of Experimental Social Psychology*, 43(4), 539-552.
- Schulz-Hardt, S., Brodbeck, F. C., Mojzisch, A., Kerschreiter, R., & Frey, D. (2006). Group decision making in hidden profile situations: Dissent as a facilitator for decision quality. *Journal of Personality and Social Psychology*, 91(6), 1080-1093.
- Schweiger, D. M., Sandberg, W. R., & Ragan, J. W. (1986). Group approaches for improving strategic decision making: A comparative analysis of dialectical inquiry, devil's advocacy, and consensus. *Academy of Management Journal*, 29(1), 51-71.
- Shaw, M. E. (1932). A comparison of individuals and small groups in the rational solution of complex problems. *The American Journal of Psychology*, 44(3), 491-504.
- Shaw, M. (1971). *Group dynamics: The psychology of small group behavior*. New York: McGraw-Hill
- Sloman, S., & Fernbach, P. (2018). *The knowledge illusion: Why we never think alone*. UK: Macmillan.
- Smoke, W. H., & Zajonc, R. B. (1962). On the reliability of group judgments and decisions. In J. Criswell, H. Solomon, & P. Suppes (Eds.), *Mathematic methods in small group process* (pp. 322-333). Stanford, CA: Stanford University Press.
- Sona-Systems. (2016). Retrieved from <http://www.sona-systems.com/>
- Spearman, C. E. (1904). 'General intelligence' objectively determined and measured. *American Journal of Psychology*, 15(2), 201-293.
- Stasser, G. (1988). Computer simulation as a research tool: The DISCUSS model of group decision making. *Journal of Experimental Social Psychology*, 24(5), 393-422.
- Stasser, G. (1999). The uncertain role of unshared information in collective choice. In L. L. Thompson, J. M. Levine & D. M. Messick (Eds.), *Shared cognition in organizations: The management of knowledge* (pp. 49-70). New York: Psychology Press.
- Stasser, G., & Abele, S. (2019). Collective choice, collaboration, and communication. *Annual Review of Psychology*, 71, 589-612.
- Stasser, G., & Titus, W. (1985). Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of Personality and Social Psychology*, 48(6), 1467-1478.

- Steiner, I. D. (1966). Models for inferring relationships between group size and potential group productivity. *Behavioral Science*, 11(4), 273-283.
- Steiner, I. D. (1972). *Group process and productivity*. New York: Academic Press.
- Steiner, I. D. (1974). Whatever happened to the group in social psychology? *Journal of Experimental Social Psychology*, 10(1), 94-108.
- Stevenson, R. J., & Over, D. E. (2001). Reasoning from uncertain premises: Effects of expertise and conversational context. *Thinking & Reasoning*, 7(4), 367-390.
- Sugimoto, M., & Kusumi, T. (2014). The effect of text continuity on spatial representation: Route versus survey perspective. *Cognitive Processing*, 15(1), 65-75.
- Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations*. New York, NY: Doubleday.
- Talmy, L. (1983). How language structures space. In H. L. J. Pick & L. P. Acredolo (Eds.), *Spatial orientation: Theory, research, and applications* (pp. 225-282). New York, NY: Plenum Press.
- Tarski, A. (1936). Der Wahrheitsbegriff in den formalisierten Sprachen. *Studia Philosophica*, 1, 261-405.
- Thomas, E. J., & Fink, C. F. (1961). Models of group problem solving. *The Journal of Abnormal and Social Psychology*, 63(1), 53-63.
- Thompson, V. A. (2009). Dual process theories: A metacognitive perspective. In J. Evans & K. Frankish (Eds.), *Two minds: Dual processes and beyond* (pp. 171-195). New York: Oxford University Press.
- Thompson, V. A., Prowse-Turner, J. A., & Pennycook, G. (2011). Intuition, reason and metacognition. *Cognitive Psychology*, 63(3), 107-140.
- Tindale, R. S., & Kameda, T. (2000). 'Social sharedness' as a unifying theme for information processing in groups. *Group Processes & Intergroup Relations*, 3(2), 123-140.
- Tindale, R. S., Nadler, J., Krebel, A., & Davis, J. H. (2004). Procedural Mechanisms and Jury Behavior. In M. A. Hogg & R. S. Tindale (Eds.), *Blackwell handbook of social psychology: Group processes* (pp. 136-164). Malden: Blackwell.
- Tindale, R. S., Smith, C. M., Dykema-Engblade, A., & Kluwe, K. (2012). Good and bad group performance: Same process – different outcomes. *Group Processes & Intergroup Relations*, 15(5), 603-618.
- Tindale, R. S., Smith, C. M., Thomas, L. S., Filkins, J., & Sheffey, S. (1996). Shared representations and asymmetric social influence processes in small groups. In E. Witte & J. H. Davis (Eds.), *Understanding group behavior: Consensual action by small groups* (Vol. 1, pp. 81–103). New York: Mahwah.
- Tindale, R. S., & Winget, J. R. (2019). Group decision-making. In M. A. Hogg & J. Cooper (Eds.) *Sage handbook of social psychology*. London: Sage Publications, Inc.

- Tröbst, S., Hardy, I., & Möller, K. (2011). Die Förderung deduktiver Schlussfolgerungen bei Grundschulkindern in naturwissenschaftlichen Kontexten. *Unterrichtswissenschaft*, 39(1), 7-20.
- Tversky, B. (2011). Spatial thought, social thought. In T. W. Schubert & A. Maass (Eds.), *Spatial dimensions of social thought* (pp. 17-38). Berlin, Boston: De Gruyter Mouton.
- Urbaniak, G. C., & Plous, S. (2013). Research Randomizer (Version 4.0) [Computer software]. Retrieved on June 22, 2013, from <http://www.randomizer.org/>
- Vajsbaheer, T., Schultheis, H., & Francis, N. K. (2018). Spatial cognition in minimally invasive surgery: A systematic review. *BMC Surgery*, 18(1), 94-109.
- Van de Ven, A. H., & Delbecq, A. L. (1972). The nominal group as a research instrument for exploratory health studies. *American Journal of Public Health*, 62(3), 337-342.
- Van de Ven, A. H., & Delbecq, A. L. (1974). The effectiveness of nominal, delphi and interacting group decision making processes. *Academy of Management Journal*, 17(4), 605-621.
- Vandierendonck, A., & De Vooght, G. (1997). Working memory constraints on linear reasoning with spatial and temporal contents. *The Quarterly Journal of Experimental Psychology: Section A*, 50(4), 803-820.
- Vandierendonck, A., Dierckx, V., & De Vooght, G. D. (2004). Mental model construction in linear reasoning: Evidence for the construction of initial annotated models. *The Quarterly Journal of Experimental Psychology*, 57(8), 1369-1391.
- Vogel, T., Wanke, M. (2016). *Attitudes and attitude change*. London: Psychology Press,
- Von Cranach, M., Ochsenein, G., & Valach, L. (1986). The group as a self-active system: Outline of a theory of group action. *European Journal of Social Psychology*, 16(3), 193-229.
- Walsh, J. P., & Maloney, N. G. (2007). Collaboration structure, communication media, and problems in scientific work teams. *Journal of Computer-Mediated Communication*, 12(2), 712-732.
- Wason, P. C. (1966). Reasoning. In B. M. Foss (Ed.), *New horizons in psychology* (pp. 135-151). Harmondsworth, UK: Penguin.
- Wegner, D. M. (1987). Transactive memory: A contemporary analysis of the group mind. In *Theories of group behavior* (pp. 185-208). Springer, New York, NY.
- Witte, E. H. (2014). Gruppe. In G. Endruweit, G. Trommsdorff & N. Burzan (Eds.), *Wörterbuch der Soziologie* (S. 158-163). Konstanz: UVK.
- Wolf, A. G., Rieger, S., & Knauff, M. (2012). The effects of source trustworthiness and inference type on human belief revision. *Thinking/ & Reasoning*, 18(4), 417-440.
- Woolley, A. W., Aggarwal, I., & Malone, T. W. (2015). Collective intelligence and group performance. *Current Directions in Psychological Science*, 24(6), 420-424.

-
- Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science*, 330(6004), 686-688.
- World Medical Association (2013). World medical association declaration of Helsinki: Ethical principles for medical research involving human subjects. *JAMA*, 310(20), 2191–2194.
- Zajonc, R. B. (1965). Social facilitation. *Science*, 149(3681), 269-274.

Anhang

Anhang A

Tabelle 17 (erster Teil)

Prädikatenlogische Schemata und Modelle für das Aufgabenmaterial von Experiment 3-5

Ein Modell Probleme						
Durchgang	1. Prämisse	2. Prämisse	3. Prämisse	4. Prämisse	Modelle	
Übung 1	A l B	C r B	D u A	E u C	A B C D E	
Problem 1	C u B	B u A	D r A	E r C	A D B C E	
Problem 2	B ü C	A ü B	D l A	E l C	D A B E C	
Problem 3	B r A	B l C	E ü C	D ü A	D E A B C	
Problem 4	B r A	B l C	E u C	D u A	A B C D E	
Zwei-Modelle determinierte Probleme						
Übung 2	C ü A	A u B	D l A	E l C	E C B D A	B E C D A
Problem 5	C r A	B r A	D u C	E u A	A B C E D	A C B E D
Problem 6	A ü B	C u A	D r A	E r C	A D B C E	A D C E B
Problem 7	B u A	A ü B	E l C	D l A	D A B E C	D A E C B
Problem 8	A l C	A l B	D ü A	E ü C	D E A B C	D E A C B

Anmerkung: Prädikate stehen für: r =rechts, l =links, u =vor (Exp3) unter (Exp 4+5), ü = hinter (Exp 4+5). Die Objekte waren in Experiment 3 und 4 verschieden farbige Blöcke / Quadrate: Orange, blau, rot, grün, gelb. In Experiment 5 waren die Terme Obst (Apfel, Birne, Mango, Kiwi, Orange) oder Werkzeug (Hammer, Säge, Meißel, Feile, Bohrer).
Tabelle wird auf der nächsten Seite fortgeführt.

Tabelle 17 (zweiter Teil)

Prädikatenlogische Schemata und Modelle für das Aufgabenmaterial von Experiment 3-5

Zwei-Modelle indetermierte Probleme						
Durchgang	1. Prämisse	2. Prämisse	3. Prämisse	4. Prämisse	Modelle	
Übung 3	C r B	A l C	E u A	D u B	A B C E D	B A C D E
Problem 9	B r A	A l C	D u B	E u C	A B C D E	A C B E D
Problem 10	A ü B	A ü C	D r B	E r C	A B D C E	A C E B D
Problem 11	C u A	B ü C	E l A	D l B	E A D B C	D B E A C
Problem 12	C l A	B l A	E ü C	D ü B	E D C B A	D E B C A

Anmerkung: Prädikate stehen für: r =rechts, l =links, u =vor (Exp3) unter (Exp 4+5), ü = hinter (Exp3) über (Exp 4+5).

Die Objekte waren in Experiment 3 und 4 verschieden farbige Blöcke / Quadrate: Orange, blau, rot, grün, gelb. In Experiment 5 waren die Terme Obst (Apfel, Birne, Mango, Kiwi, Orange) oder Werkzeug (Hammer, Säge, Meißel, Feile, Bohrer).

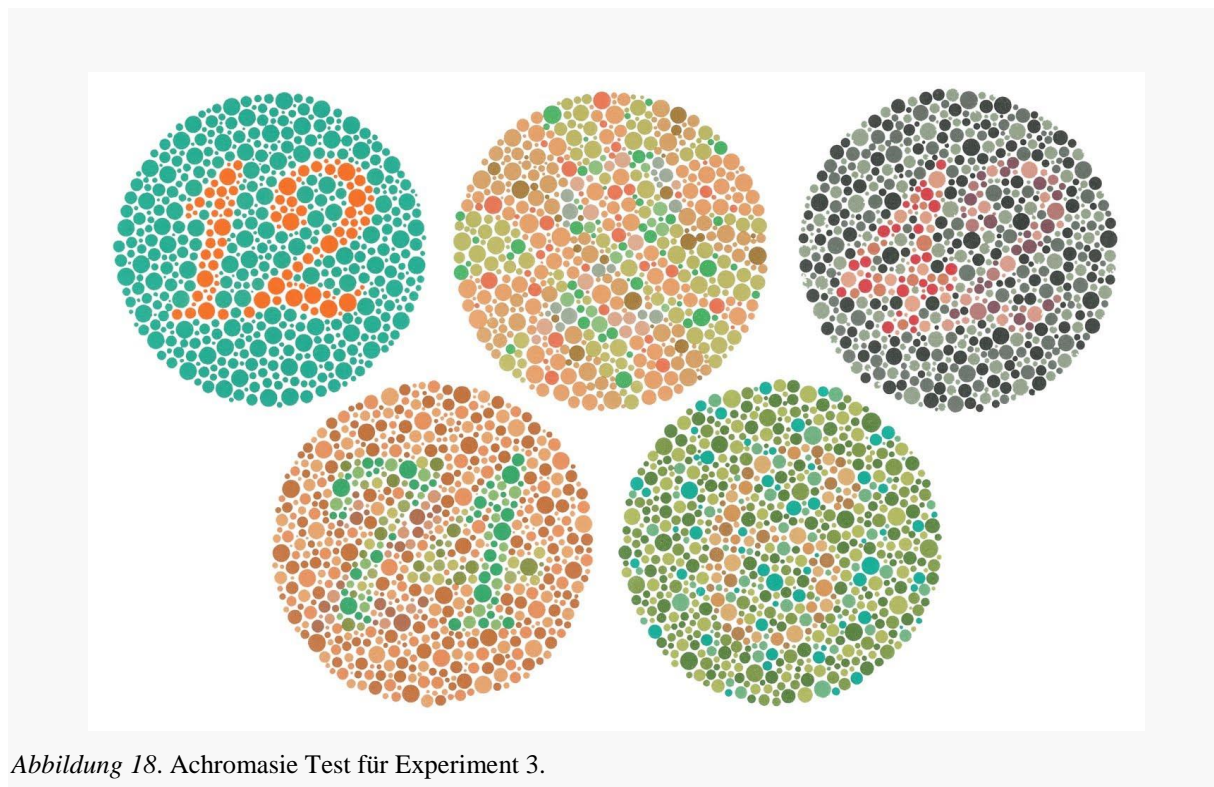
Anhang B

Abbildung 18. Achromasie Test für Experiment 3.

Anhang C

Tabelle 18

Mittleren Bewertungen der Vertrauenswürdigkeit aus der Vorstudie (links) und dem Hauptexperiment (rechts) für das Experiment 5

<i>M</i>	Vorstudie	Vertrauens- würdigkeit	Hauptexperiment	<i>M</i>
5.89	Feuerwehrmann	Hoch	Feuerwehrmann	5.93
5.74	Tierarzt		Psychologe	5.80
5.72	Richter		Tierarzt	5.73
5.72	Pilot		Richter	5.47
5.40	Psychologe		Pilot	5.33
5.35	Polizist		Polizist	4.67
5.22	Entwicklungshelfer		Entwicklungshelfer	4.67
4.68	Metzger		Mittel	Jäger
4.67	Fotograf	Fotograf		4.60
4.51	Student	Metzger		4.60
4.49	Sachbearbeiter	Student		4.53
4.48	Jäger	Radiomoderator		4.47
4.39	Friseur	Sachbearbeiter		4.33
4.26	Berater	Berater		4.33
4.19	Radiomoderator	Friseur		4.07
3.51	Minister	Niedrig	Minister	3.73
3.42	Bundestagsabgeordneter		Spion	3.60
3.24	Spion		Börsenmakler	3.33
2.95	Börsenmakler		Bundestagsabgeordneter	3.13
2.86	Versicherungsvertreter		Versicherungsvertreter	2.80
2.85	Immobilienmakler		Immobilienmakler	2.60
2.66	Gebrauchtwagenhändler		Gebrauchtwagenhändler	2.53

Erklärung

Ich erkläre: Ich habe die vorgelegte Dissertation selbstständig und nur mit den Hilfen angefertigt, die ich in der Dissertation angegeben habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten oder nicht veröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht.

Gießen, den 13.04.2020

Andreas Reis