

# Kausales Denken, Bayes-Netze und die Markov-Bedingung

DISSERTATION

zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades

“Doctor rerum naturalium“

der Georg-August-Universität Göttingen

vorgelegt von

Ralf Mayrhofer

aus Magdeburg

Göttingen, im Januar 2009

Referent: Prof. Dr. Michael R. Waldmann

Korreferent: Prof. Dr. Matthias Nückles

Tag der mündlichen Prüfung: 11.02.2009

*If the brain were so simple we could understand it,  
we would be so simple we couldn't.*

Lyall Watson

---

## Inhaltsverzeichnis

1	Einleitung.....	1
2	Theoretischer und empirischer Hintergrund.....	6
2.1	Einfache psychologische Modelle kausalen Lernens .....	6
2.1.1	<i>Assoziationstheoretische Ansätze</i> .....	6
2.1.2	<i>Kovariationsbasierte Ansätze</i> .....	7
2.1.3	<i>Theorie kausaler Mechanismen</i> .....	9
2.2	Kausale Bayes-Netze als psychologische Theorie kausalen Denkens ....	10
2.2.1	<i>Einführung in die kausale Bayes-Netz-Theorie</i> .....	10
2.2.2	<i>Kausale Bayes-Netze als psychologische Theorie</i> .....	13
2.2.3	<i>Das Causal-Support-Modell (Griffiths &amp; Tenenbaum, 2005)</i> .....	18
2.3	Empirische Untersuchungen zur Markov-Bedingung .....	21
2.4	Kausale Prozesse, Fehlerstrukturen und die Markov-Bedingung .....	25
2.5	Zusammenfassung.....	28
3	Basisexperimente .....	30
3.1	Experiment 1 .....	33
3.1.1	<i>Methode</i> .....	33
3.1.2	<i>Ergebnisse</i> .....	39
3.1.3	<i>Diskussion</i> .....	41
3.2	Experiment 2 .....	42
3.2.1	<i>Methode</i> .....	42
3.2.2	<i>Ergebnisse</i> .....	44
3.2.3	<i>Diskussion</i> .....	46
3.3	Experiment 3 .....	47
3.3.1	<i>Methode</i> .....	48
3.3.2	<i>Ergebnisse</i> .....	51

---

3.3.3	<i>Diskussion</i> .....	54
3.4	Zusammenfassung.....	54
4	Modellierung.....	56
4.1	Modellierungsidee.....	57
4.2	Ableitung des Basismodells.....	60
4.3	Modellvorhersagen.....	63
4.4	Sensitivitätsanalysen.....	66
4.4.1	<i>Abhängigkeit von der Basisrate der Effekte (<math>b_E</math>)</i> .....	67
4.4.2	<i>Abhängigkeit der Basisrate der Ursache (<math>b_C</math>)</i> .....	68
4.4.3	<i>Abhängigkeit von der Kausalstärke von C (<math>w_C</math>)</i> .....	68
4.4.4	<i>Abhängigkeit von der Basisrate der Fehlerursache (<math>b_{PN}</math>)</i> .....	70
4.5	Fitanalyse.....	73
4.6	Zusammenfassung und Diskussion.....	75
5	Modelltests.....	77
5.1	Schließen im Fall zweier kausal aktiver Zustände der Ursache.....	77
5.1.1	<i>Modellierung</i> .....	78
5.1.2	<i>Experiment 4: Überblick</i> .....	80
5.1.3	<i>Experiment 4: Methode</i> .....	81
5.1.4	<i>Experiment 4: Ergebnisse</i> .....	86
5.1.5	<i>Diskussion</i> .....	88
5.2	Schließen in kausalen Ketten.....	88
5.2.1	<i>Modellierung</i> .....	89
5.2.2	<i>Experiment 5: Überblick</i> .....	92
5.2.3	<i>Experiment 5: Methode</i> .....	93
5.2.4	<i>Experiment 5: Ergebnisse</i> .....	96
5.2.5	<i>Diskussion</i> .....	98

---

5.3	Schließen im Fall weiterer beobachteter präventiver Ursachen .....	99
5.3.1	<i>Modellierung</i> .....	99
5.3.2	<i>Experiment 6: Überblick</i> .....	103
5.3.3	<i>Experiment 6: Methode</i> .....	104
5.3.4	<i>Experiment 6: Ergebnisse</i> .....	108
5.3.5	<i>Experiment 6: Diskussion</i> .....	111
5.3.6	<i>Experiment 7: Überblick</i> .....	111
5.3.7	<i>Experiment 7: Methode</i> .....	111
5.3.8	<i>Experiment 7: Ergebnisse</i> .....	113
5.3.9	<i>Diskussion</i> .....	114
5.4	Zusammenfassung.....	115
6	Modellerweiterung: Merkmalsbasierte Inferenz .....	118
6.1	Idee und Überblick .....	118
6.2	Modellierung .....	120
6.2.1	<i>Ableitung der Zielinferenz</i> .....	120
6.2.2	<i>Zielinferenz gegeben einer Partitionierung <math>z</math></i> .....	122
6.2.3	<i>Wahrscheinlichkeit einer Partitionierung <math>z</math> gegeben <math>F</math></i> .....	124
6.2.4	<i>Modellvorhersagen</i> .....	127
6.3	Experiment 8 .....	129
6.3.1	<i>Methode</i> .....	130
6.3.2	<i>Ergebnisse</i> .....	134
6.4	Diskussion .....	137
6.5	Zusammenfassung.....	140
7	Gesamtdiskussion und Ausblick.....	142
7.1	Diskussion der Befunde .....	142
7.2	Ausblick.....	150

---

8	Zusammenfassung.....	153
9	Literaturverzeichnis .....	154
	Anhang A – Varianzanalytische Ergebnisse von Experiment 1 .....	167
	Anhang B – Varianzanalytische Ergebnisse von Experiment 2 .....	169
	Anhang C – Varianzanalytische Ergebnisse von Experiment 3 .....	171
	Anhang D – Ausführliche Ableitung des Basismodells .....	173
	Anhang E – Varianzanalytische Ergebnisse von Experiment 4 .....	176
	Anhang F – Herleitung des Modell für <i>Causal-Chain</i> -Strukturen.....	178
	Anhang G – Varianzanalytische Ergebnisse – Experiment 5.....	180
	Anhang H – Herleitung für den Fall weiterer Fehlerquellen.....	182
	Anhang I – Varianzanalytische Ergebnisse – Experiment 6 .....	186
	Anhang J – Varianzanalytische Ergebnisse – Experiment 7 .....	189
	Anhang K – Ausführliche Herleitung des erweiterten Modells .....	190
	Anhang L – Detaillierte Vorhersage des erweiterten Modells.....	194
	Anhang M – Detaillierte Ergebnisse von Experiment 8.....	196
	Anhang N – Varianzanalytische Ergebnisse für Experiment 8 .....	198
	<i>Lebenslauf</i> .....	199

## 1 Einleitung

Die Fähigkeit des Menschen, kausale Beziehungen in der Welt zu entdecken und das Wissen um diese nutzbar zu machen, um zukünftige Ereignisse vorherzusagen, vergangene Ereignisse zu erklären und Handlungen zur Manipulation solcher Ereignisse zu planen, ist eine zentrale Kompetenz, um in seiner Umwelt erfolgreich agieren zu können. Dazu gehört nicht nur das Lernen einfacher Ursache-Effekte-Relationen, sondern auch komplexeres, vernetztes Kausalwissen, das mehr als zwei Variablen umfasst.

Die Theorie der kausalen Bayes-Netze, wie sie in den vergangenen zwanzig Jahren insbesondere im Bereich der Informatik entwickelt worden war (siehe Pearl, 1988, 2000; Heckerman, 1999; Spirtes, Glymour & Scheines, 2000), ist seit einigen Jahren daher in der Psychologie ein populärer Ansatz, um menschliches Kausaldenken und -lernen in solch komplexeren, über den Eine-Ursache-ein-Effekt-Fall hinausgehenden Situationen zu beschreiben (siehe u.a. Glymour, 2001, 2003; Gopnik, Glymour, Sobel, Schulz, Kushnir & Danks, 2004; Sloman, 2005; Steyvers, Tenenbaum, Wagenmakers & Blum, 2003; Waldmann & Martignon, 1998). Zentrale Annahme dieses Ansatzes ist die Markov-Bedingung, nach der eine Variable konditionalisiert auf ihre direkten Ursachen unabhängig von allen anderen Variablen des Systems ist, sofern diese nicht (direkte oder indirekte) Effekte eben jener Variable sind. In einer einfachen *Common-Cause*-Struktur mit einer Ursache und mehreren Effekten bedeutet dies, dass die auf die Ursache bedingte Wahrscheinlichkeit des Auftretens eines Effekts unabhängig von der An- oder Abwesenheit der anderen Effekte dieser Ursache ist.

Die Annahme der Markov-Bedingung ermöglicht damit einfache Inferenzen selbst in komplexesten kausalen Strukturen und bildet die Grundlage für viele Algorithmen zum Lernen kausaler Netze im Bereich der Informatik (siehe Heckerman, 1999; Spirtes et al., 2000). Ohne die Markov-Bedingung wären für die Inferenz hinsichtlich einer Variablen des kausalen Systems der Status aller anderen Variablen zu berücksichtigen. Daher scheint die Markov-Bedingung nicht nur aus theoretischer Perspektive sinnvoll und notwendig, sondern auf den ersten Blick im Grundsatz auch psychologisch plausibel.



---

Neuere Forschung, in der die Intuitionen von Probanden hinsichtlich dieser Unabhängigkeitsannahme untersucht wurde, nährt jedoch Zweifel an der Markov-Bedingung als Teil einer psychologischen Theorie des kausalen Denkens (siehe Rehder & Burnett, 2005; Waldmann, Cheng, Hagmayer & Blaisdell, 2008; Walsh & Sloman, 2007). So hängt die Beurteilung der Anwesenheit eines Effektes in einer *Common-Cause*-Struktur gegeben der An- oder Abwesenheit seiner Ursache sehr wohl und gleichsam auch sehr deutlich vom Status der weiteren Effekte ab. Diese „Markov-Verletzung“ zeigte sich dabei in allen untersuchten kausalen Domänen (Rehder, 2006; Rehder & Burnett, 2005).

Die interessante Frage ist nun, wie dieses augenscheinlich nicht normative Verhalten zu bewerten ist. Eine mögliche Interpretation ist, dass menschliches Kausaldenken nicht rational im normativen Sinne ist. Kausale Bayes-Netze sind demnach zu komplex, um die einfachen, nicht-normativen Strategien, wie sie möglicherweise von Menschen genutzt werden, abzubilden (siehe dazu z.B. Waldmann et al., 2008; Walsh & Sloman, 2007). Eine andere Interpretationsmöglichkeit ist, dass Menschen viel mehr Informationen und Wissen bei ihren Inferenzen berücksichtigen, als eine einfache Analyse von kausalen Situationen auf Bayes-Netz-Ebene suggeriert (siehe u.a. Lu, Yuille, Liljeholm, Cheng & Holyoak, 2008; Mayrhofer, Goodman, Waldmann & Tenenbaum, 2008). Demnach sind Bayes-Netze nicht zu komplex, sondern zu einfach, um menschliche Kausalinferenzen adäquat abzubilden und zu beschreiben. Im Sinne dieser Interpretationsmöglichkeit müssen sie entsprechend erweitert werden, um als Teil einer psychologischen Theorie kausalen Denkens und Lernens eingesetzt werden zu können.

Letztere Idee basiert dabei auf der von Marr (1982) und in ähnlicher Weise von Anderson (1990) eingeführten Unterscheidung zwischen verschiedenen Ebenen der theoretischen Betrachtung: der computationalen Ebene, der Repräsentations- und algorithmischen Ebene sowie der Ebene der (Hardware-) Implementierung. Auf der hierfür interessierenden computationalen Ebene geht es dabei vor allem um die Frage, welches Problem das kognitive System z.B. in einer experimentellen Situation eigentlich zu lösen versucht und auf welchen Repräsentationen es dabei operiert. Die Rationalität der Inferenzen wird dabei  $\alpha$

---

*priori* angenommen (Anderson, 1990), im Sinne einer optimalen Adaptation des menschlichen Gehirns an seine Umwelt.<sup>1</sup> Weicht beobachtetes Verhalten von postulierten Normen (= computationale Theorie) ab, so dokumentiert dies die Unangepasstheit der Normen, an denen wir das menschliche Denken messen, und nicht die Unangemessenheit des menschlichen Denkens. Ziel muss es folglich sein, computationale Theorien zu entwickeln, die menschliches Verhalten beschreiben und vorhersagen, sich dabei aber gleichzeitig in der Form der Beschreibung auf der Basis normativer Axiome (wie z.B. Wahrscheinlichkeitskalkül) bewegen und sich daraus in konsistenter Weise ableiten.<sup>2</sup>

Ziel der vorliegenden Arbeit ist es, dieses Konzept auf menschliches Kausaldenken anzuwenden und dieses auf computationaler Ebene in Situationen zu untersuchen, in denen es sich nach den Untersuchungen von Rehder und Burnett (2005) als „Markov verletzend“ herausgestellt hat. Die Arbeit gliedert sich dabei wie folgt:

In *Kapitel 2* werden als Ausgangspunkt die Literatur zum kausalen Lernen und Denken überblicksartig dargestellt und eine Einführung in die aktuellere Forschung im Bereich der Bayes-Netze als psychologische Theorien kausalen Denkens und die damit als zentrale Annahme eng verbundene Markov-Bedingung gegeben. Dabei werden insbesondere die Arbeiten von Rehder und Burnett (2005) genauer vorgestellt, die die Abweichungen von dieser zentralen Annahme („Markov-Verletzungen“) empirisch untersucht und dokumentiert haben. Die Studien dienen sodann als Motivation für die Entwicklung der Idee, dass Probanden bei Kausalinferenzen systematisch weiteres Wissen, insbesondere um die zugrunde liegenden kausalen Prozesse, miteinbeziehen und dieses sich vor allem auf die repräsentierte Fehlerstruktur auswirkt.

---

<sup>1</sup> Diese Sichtweise ist dabei vor allem aus der Perspektive der Künstlichen-Intelligenz-Forschung motiviert, da es bis heute nicht gelungen ist, eine Maschine zu konstruieren, die den Leistungen des menschlichen Gehirns, z.B. im Bereich der visuellen Wahrnehmung oder der autonomen Handlungsplanung und -steuerung in komplexen Situationen, auch nur im Entferntesten nahe kommt. Aus dieser Perspektive erscheint es daher vermessen, menschliches Verhalten deshalb als irrational zu deklarieren, weil es von „normativen“ Modellen abweicht, die sich in realen Situationen aufgrund ihrer simplifizierenden Annahmen überhaupt nicht implementieren lassen. In diesem Sinne gilt mithin das menschliche Verhalten und Denken als Messlatte, die es (vorerst) zu erreichen gilt und nicht umgekehrt.

<sup>2</sup> Für eine ausführliche Debatte dieses Ansatzes siehe u.a. Brighton & Gigerenzer, 2008; Chater und Oaksford (2008a, 2008b), Danks (2008), Oaksford und Chater (2007), Shultz (2007), Sloman und Fernbach (2008) sowie Waldmann et al. (2008).

---

Diese Idee wird in *Kapitel 3* in drei Experimenten (Basisexperimente) getestet und die Ergebnisse dargestellt. Hierzu wird das „*Mind-Reading-Alien*“-Paradigma, wie es von Steyvers et al. (2003) eingeführt wurde, mit der Aufgabe von Rehder und Burnett (2005) verknüpft: In einer *Common-Cause*-Struktur mit drei Effekten sind von den Probanden Schätzungen über die Anwesenheit eines nicht beobachteten (aber beobachtbaren) Effekts gegeben des Status der anderen Variablen abzugeben. Dabei wird das Wissen um die zugrunde liegenden Kausalprozesse (Gedankenlesen vs. Gedankensenden) manipuliert und gezeigt, dass die Stärke der gefundenen Markov-Verletzung über diese Manipulationen systematisch variiert und dass dies auf die jeweils unterschiedlichen Annahmen der Probanden hinsichtlich der durch die zugrunde liegenden Kausalprozesse implizierten Fehlerstruktur zurückzuführen ist.

Darauf aufbauend wird in *Kapitel 4* zur Erklärung des Phänomens ein bayesianisches Modell abgeleitet, welches klassische Bayes-Netze um eine ursacheseitige Fehlerkomponente erweitert, deren Stärke von den Annahmen hinsichtlich der zugrunde liegenden Kausalprozesse abhängig ist. Der Schlussprozess wird dabei als adaptive Fehlerattribution modelliert. Darauf aufbauend werden aus dem Modell Vorhersagen abgeleitet und die qualitativ-strukturellen Modellvorhersagen mit den Befunden aus den Basisexperimenten verglichen. Des Weiteren wird in einer Fitanalyse sowohl auf qualitativer wie auch quantitativer Ebene überprüft, wie sehr die qualitativ-strukturelle Modellvorhersage von der Wahl der freien Modellparameter abhängig ist.

In *Kapitel 5* werden aus dem Modell für drei Anwendungsbereiche weitere, neue Vorhersagen abgeleitet und jeweils empirisch geprüft, um das Modell explizit zu testen. So wird das Modell in Abschnitt 5.1 auf eine *Common-Cause*-Struktur angewendet, in der in Abweichung zur üblichen Interpretation beide Zustände der gemeinsamen Ursache kausal aktiv sind. In Abschnitt 5.2. wird die dem Modell zugrunde liegende Idee der Einführung einer ursacheseitigen Fehlerquelle auf eine kausale Kette erweitert und entsprechende Vorhersagen generiert. In Abschnitt 5.3 wird der *Common-Cause*-Struktur, wie sie in den Basisexperimenten Verwendung fand und wie sie Ausgangspunkt für die Modellierung war, eine beobachtete präventive Ursache hinzugefügt, die mit der modellendogenen

ursacheseitigen Fehlerquelle im Hinblick auf die adaptive Fehlerattribution konkurriert.

Das Modell wird sodann in *Kapitel 6* im Hinblick auf die Frage beleuchtet, inwieweit die Inferenzen in kausalen Systemen auch von den weiteren, auf den ersten Blick nicht kausalen Eigenschaften, insbesondere den Merkmalen (Features) der involvierten Objekte und damit von möglichen Kategorisierungen der Variablen abhängig sind. Unter der Annahme, dass die durch die ursacheseitig eingeführte Fehlerquelle implizierten konditionalen Abhängigkeiten nur zwischen Effekten bestehen, die einem Cluster zugeordnet sind, und dass solche Cluster im Hinblick auf kausal relevante Merkmale der involvierten Objekte möglichst homogen sind, wird eine Erweiterung des in Kapitel 4 vorgestellten Modells entwickelt, welches die Zielinferenz über alle möglichen Partitionierungen der Effekte des Systems bestimmt und über die Unsicherheit bezüglich der Clusterzugehörigkeit integriert. Aus diesem erweiterten Modell werden dann entsprechend ebenfalls Vorhersagen abgeleitet und in einem Experiment getestet.

Die Arbeit schließt in *Kapitel 7* mit einer Gesamtdiskussion und einem Ausblick. Zum einen sollen hierbei das Modell und die durchgeführten Experimente in ihrer Gesamtschau gewürdigt werden. Zum anderen wird aufbauend auf den in der Arbeit entwickelten Ideen ein umfassenderes, computationales Modell des Kausaldenkens skizziert, das zwischen der rein statistischen Ebene der Ereignisse und einer kausalen Hintergrundebene, in der die involvierten Objekte und auch Mechanismen repräsentiert sind, unterscheidet.

## 2 Theoretischer und empirischer Hintergrund

Im vorliegenden Kapitel werden als Ausgangspunkt die Literatur zum kausalen Lernen und Denken überblicksartig dargestellt und eine Einführung in die aktuellere Forschung im Bereich der Bayes-Netze als psychologische Theorien kausalen Denkens und die damit als zentrale Annahme eng verbundene Markov-Bedingung gegeben. Dazu soll als erstes ein kurzer Überblick über einfache psychologische Modelle des Kausallernens, wie assoziations-theoretische Ansätze, kovariationsbasierte Ansätze ( $\Delta P$ -Regel, *Causal Power*) und mechanismusbasierte Ansätze gegeben werden und daran anschließend eine Einführung in die Kausale-Bayes-Netz-Theorie und ihrer Anwendung in der Psychologie erfolgen. Etwas genauer wird auf das *Causal-Support*-Modell von Griffiths und Tenenbaum (2005) eingegangen, da dieses kausale Bayes-Netze mit den psychologischen Ansätzen verbindet und einige wichtige Grundlagen für das später in Kapitel 4 entwickelte Modell legt.

Des Weiteren werden die empirischen Arbeiten von Rehder und Burnett (2005) genauer vorgestellt, die Abweichungen von der Markov-Bedingung, die Grundlage für die auf der Bayes-Netz-Theorie basierenden Modelle ist, empirisch untersucht und dokumentiert haben. Die Studien dienen sodann als Motivation für die Entwicklung der Idee, dass Probanden bei Kausalinferenzen systematisch weiteres Wissen, insbesondere über die zugrunde liegenden kausalen Prozesse und die daraus resultierende Fehlerstruktur miteinbeziehen.

### 2.1 Einfache psychologische Modelle kausalen Lernens

Im Fokus früherer psychologischer Modelle stand vor allem die Frage, wie eine Kausalrelation zwischen zwei betrachteten Ereignissen gelernt werden kann (für einen Überblick siehe z.B. Hagmayer & Waldmann, 2006, sowie Perales & Catena, 2006). Dabei kann zwischen den assoziations-theoretischen Ansätzen, den kovariationsbasierten Ansätzen sowie den mechanismusbasierten Ansätzen unterschieden werden.

#### 2.1.1 Assoziations-theoretische Ansätze

Assoziations-theoretische Lernmodelle in der Psychologie haben eine lange Tradition. Entwickelt wurden diese ursprünglich, um den Erwerb konditionierter Reak-

tionen (klassisches Konditionieren) sowie das Lernen von einfachen Verhaltensweisen (instrumentelles Konditionieren) zu erklären und zu beschreiben. Diese Modelle, wie zum Beispiel das Rescorla-Wagner-Modell (Rescorla & Wagner, 1972), beschreiben dabei, wie zwischen einem Hinweisreiz (*cue*, z.B. einem Ton) und dessen potentieller Folge (*outcome*, z.B. Futter) eine Assoziation gebildet wird – i.d.R. indem die stetige Paarung von *Cue* und *Outcome* die entsprechende assoziative Stärke erhöht. In Folge sind diese Modelle auch auf höhere kognitive Prozesse, wie Kausallernen, angewendet worden (siehe u.a. Dickinson, 2001; Shanks, 2007). Demnach sei das Lernen von Kausalrelationen nichts anderes als das Lernen von *Cue-Outcome*-Relationen, die Einschätzung der Stärke einer Ursache-Effekt-Relation entspricht damit deren assoziativen Stärke.

Die assoziativen Ansätze sind damit allerdings nicht sensitiv für die unterschiedlichen Rollen, denen Ursache und Effekt beizumessen sind (siehe u.a. Waldmann, 1996; Waldmann & Holyoak, 1992), vor allem nicht im Hinblick auf die fundamentale Asymmetrie einer Kausalrelation: Ursachen produzieren ihre Effekte und nicht umgekehrt.

### 2.1.2 Kovariationsbasierte Ansätze

Anders als die eben skizzierten Ansätze, die Kausalrelationen lediglich als gelernte Assoziationen begreifen, gehen die kovariationsbasierten Modelle einen Schritt weiter, indem sie die unterschiedliche Rolle von Ursache und Effekt aufgreifen: Ursachen erhöhen die Wahrscheinlichkeit der Anwesenheit ihrer Effekte, d.h. ist die Variable *C* eine Ursache und die Variable *E* deren Effekt, dann sollte gelten:  $P(E^+|C^+) > P(E^+|C^-)$ .<sup>3</sup> Kausalrelationen werden damit auf der Basis der Häufigkeiten des gemeinsamen Auftretens (und Nichtauftretens) der beiden betrachteten Ereignisse *C* und *E* gelernt<sup>4</sup>.

Das einfachste und älteste Modell ist die  $\Delta P$ -Regel<sup>5</sup>, nach der mit Bezug auf genannte Ungleichung die Einschätzung der Stärke einer Kausalrelation schlicht monotone Funktion der Kontingenz  $\Delta P = P(E^+|C^+) - P(E^+|C^-)$  ist (siehe u.a. Allan & Jenkins, 1980; Jenkins & Ward, 1965). Je stärker die Anwesenheit einer

<sup>3</sup> Die Kurzform  $E^+$  und  $C^+$  stehen hier und im Folgenden für die jeweilige Anwesenheit ( $E=1$  und  $C=1$ ) und  $E^-$  und  $C^-$  für die jeweilige Abwesenheit der Variablen ( $E=0$  und  $C=0$ ).

<sup>4</sup> Diese Häufigkeiten werden dabei in der Regel in einer 2x2-Kontingenztafel dargestellt.

<sup>5</sup> Für einen ähnlichen Ansatz siehe White (2002, 2004, 2008).

potentiellen Ursache im Vergleich zu ihrer Abwesenheit die Wahrscheinlichkeit eines potentiellen Effekts erhöht, desto stärker ist der wahrgenommene kausale Einfluss der Ursache auf den Effekt.<sup>6</sup>  $\Delta P$  ist damit ein Maß der Änderung der Wahrscheinlichkeit der Anwesenheit eines Effekts, wenn die Ursache hinzukommt. Interessanterweise konnte auf theoretischer Ebene gezeigt werden, dass die assoziative Stärke, die einer Kausalrelation im Rescorla-Wagner-Modell zukommt (siehe Abschnitt 2.1.1), mit zunehmender Anzahl an Lerntrials gegen  $\Delta P$  konvergiert (Danks, 2003).

Gegen die  $\Delta P$ -Regel ist vor allem von Cheng (1997; siehe auch Cheng, Novick, Liljeholm & Ford, 2007; Novick & Cheng, 2004; Wu & Cheng, 1999) eingewendet worden, dass  $\Delta P$  ein reines Maß der Kovariation zweier Ereignisse darstellt und daher gerade nicht deren kausale Verbindung repräsentiert. In Abgrenzung dazu entwickelte Cheng die *Power-PC-Theorie*<sup>7</sup>, in der sie eine Größe namens *Causal Power* ( $p_c$ ) herleitet, für die Menschen in ihren Kausalurteilen eigentlich sensitiv sind. *Causal Power* ist dabei die Wahrscheinlichkeit, mit der eine Ursache ihren Effekt produziert, und damit ein Maß ihrer kausalen Stärke. Dies lässt sich auch kontrafaktisch beschreiben: *Causal Power* ist mithin die Wahrscheinlichkeit dafür, dass in einem Fall, in dem die Ursache abwesend war und der Effekt nicht auftrat, der Effekt aufgetreten wäre, wenn die Ursache anwesend gewesen wäre. Die Idee berücksichtigt damit explizit, dass Effekte auch durch weitere (Hintergrund-)Ursachen unabhängig von der betrachteten Ursache hervorgerufen werden können und dass die Einschätzung der kausalen Stärke einer Ursache von diesen Hintergrundursachen nicht beeinflusst werden sollte. Die Wahrscheinlichkeitsänderung hinsichtlich der Anwesenheit des Effekts durch die Einführung der Ursache (=  $\Delta P$ ) muss also relativiert werden an dem Spielraum, der durch die nicht durch  $C$  verursachte Anwesenheit von  $E$  nach oben hin überhaupt noch verbleibt [=  $1 - P(E^+|C^-)$ ]:

$$p_c = \frac{\Delta P}{1 - P(E^+|C^-)} \quad (1)$$

<sup>6</sup> Das Maß  $\Delta P$  kann natürlich auch negativ sein. In diesem Fall wird die Ursache als präventiv angesehen (sie vermindert also die Auftretenswahrscheinlichkeit des Effekts).

<sup>7</sup> Das „PC“ steht für *probabilistic contrast*, da die *Power-PC-Theorie* auf ihrem Vorgänger, dem *Probabilistic-Contrast-Modell*, beruht (siehe Cheng & Novick, 1990).

Ist also zum Beispiel der betrachtete Effekt in 80% der Fälle ohne die betrachtete Ursache anwesend und steigt seine Anwesenheit mit Hinzutreten der Ursache auf 90% ( $\Delta P = 0.1$ ), dann ist  $p_C = 0.5$ ; ist der betrachtete Effekt bei Abwesenheit der Ursache nie anwesend, bei deren Anwesenheit aber in 10% der Fälle anwesend ( $\Delta P = 0.1$ ), dann ist  $p_C = 0.1$ , auch wenn  $\Delta P$  in beiden Fällen identisch ist. Ist der Effekt im letzten Fall bei Anwesenheit der Ursache zu 50% anwesend ( $\Delta P = 0.5$ ), dann ist  $p_C = 0.5$ , also genau so groß wie im ersten Fall. *Causal Power* ist damit Eigenschaft der Ursache, die nicht vom Kontext – also der Häufigkeit von Hintergrundursachen – abhängig ist. Diverse empirische Studien zeigen, dass Probanden für dieses abstrakte kausale Maß sensitiv sind (Buehner, 2005; Buehner & Cheng, 2003; siehe aber für andere Befunde Collins & Shanks, 2006; Liljeholm & Cheng, in press; Lober & Shanks, 2000; Perales & Shanks, 2007)<sup>8</sup>.

### 2.1.3 Theorie kausaler Mechanismen

In Abgrenzung zu den kovariationsbasierten Ansätzen wurde die Theorie der kausalen Mechanismen entwickelt (Ahn & Bailenson, 1996; Ahn & Kalish, 2000; Ahn, Kalish, Medin & Gelman, 1995; siehe auch Koslowski, Okagaki, Lorenz & Umbach, 1989; Luhmann, 2005; Luhmann & Ahn, 2005; Perales, Catena, Maldonado & Cándido, 2007; Wolff, 2007; Wolff & Song, 2003). Die Vertreter dieses Ansatzes bringen zum einen hervor, dass kovariationsbasierte Ansätze die eigentlichen Fragen, also was verursacht was und wie wird etwas verursacht, nicht beantworten können. Die Einstufung der beobachteten Variablen als potentielle Ursache und als potentiellen Effekt ist *a priori* (siehe Cheng, 1997). *Causal Power* wie auch  $\Delta P$  lassen sich auch für die umgekehrte Richtung der betrachteten Relation berechnen. Um zu entscheiden, ob die Ursache wirklich kausal ist, müsse man weiteres Wissen, z.B. um potentielle kausale Mechanismen, berücksichtigen.<sup>9</sup> Zum anderen wird aus der Perspektive dieses Ansatzes eingewendet, dass die Einschätzung der Kausalstärke zweier Ereignisse (hier: Ursache und Effekt)

<sup>8</sup> Für theoretische Entgegnungen bezüglich der zugrunde liegenden Annahmen siehe u.a. Luhmann (2005), Luhmann und Ahn (2005) sowie White (2005).

<sup>9</sup> Eine Antwort darauf, wie man dieses Wissen erwirbt, geben die Vertreter des Ansatzes indes nicht.



deren Unabhängigkeit von Drittvariablen voraussetzt.<sup>10</sup> Diese Unabhängigkeit könne man aber nur dadurch testen, dass man alle potentiellen und damit im Zweifel unendlich viele Variablen betrachtet (Ahn & Kalish, 2000). Das Wissen um den potentiellen kausalen Mechanismus, der Ursache und Effekt verbindet, sei damit essentiell, um entscheiden zu können, welche Faktoren möglicherweise konfundierend wirken und daher kontrolliert werden müssen.

Des Weiteren konnten Ahn et al. (1995) zeigen, dass Probanden, wenn diese erklären sollten, was einen bestimmten Effekt hervorgebracht hat, zu mechanistischen Erklärungen griffen und den Effekt gerade nicht als Folge einer hohen Kovariation mit einer anderen Variable beschrieben, wie kovariationsbasierte Ansätze nahe legen. Danks (2005, siehe auch 2007) hat darauf aufbauend und in Abgrenzung zu Newsome (2003) vorgeschlagen, dass kausale Bayes-Netze die kovariationsbasierten und die mechanismusbasierten Ansätze vereinigen könnten, da Bayes-Netze neben den statistischen Eigenschaften von Ursache-Effekt-Relationen auch Informationen über die weiteren, potentiell relevanten Variablen enthalten und eine datenbasierte Entscheidung darüber, welche Variablen Ursachen und welche Effekte sind, ermöglichen.

## 2.2 Kausale Bayes-Netze als psychologische Theorie kausalen Denkens

### 2.2.1 Einführung in die kausale Bayes-Netz-Theorie

Kausale Bayes-Netze sind dem Grunde nach direkte azyklische Graphen, deren Relationen zum einen kausal interpretiert und über die zum anderen eine Wahrscheinlichkeitsfunktion definiert ist, die bestimmten Eigenschaften genügt (siehe u.a. Glymour, 2001, 2003; Pearl, 2000; Spirtes et al., 2000; Woodward, 2003).

Formal gesehen besteht ein Bayes-Netz damit aus einer Menge von Variablen  $X_1$  bis  $X_n$ , die sowohl diskret als auch kontinuierlich sein können<sup>11</sup>. Diese Variablen sind durch gerichtete Kanten (kausale Links, Kausalpfeile) verbunden, die

<sup>10</sup> Dieses Problem versucht man z.B. innerhalb der Kovariationsansätze dadurch zu lösen, dass man die eigentliche Inferenz unter Berücksichtigung bekannter konfundierender Ursachen ableitet. So wurde z.B. die *konditionale  $\Delta P$ -Regel* entwickelt, die es erlaubt, bei der Schätzung des kausalen Einflusses andere potentielle Einflussfaktoren zu kontrollieren, indem auf sie konditionalisiert wird (siehe u.a. Cartwright, 1983; Goedert & Spellman, 2005; Spellman, 1996; Waldmann, 1996).

<sup>11</sup> Derzeit werden im Bereich der psychologischen Forschung aus Vereinfachungsgründen praktisch nur binäre Variablen eingesetzt.

für jede Relation die kausale Rolle der beiden Variablen definieren: Die Variable, von der die Kante ausgeht, ist die Ursache. Die Variable, auf die die Kante zuläuft, ist ihr Effekt. Die Links selber stellen dabei zugrunde liegende, gedachte kausale Mechanismen dar, die für die resultierenden statistischen Abhängigkeiten verantwortlich sind (siehe u.a. Woodward, 2002, 2003). Sie können deterministischer oder probabilistischer Natur sein.

Die Wirkung dieser Mechanismen ist lokal und modular gedacht (siehe Pearl, 2000; Reichenbach, 1956; Steyvers et al., 2003). Der Zustand einer Variablen ist damit kausal gesehen nur vom Zustand seiner Ursachen (Eltern) abhängig. Es resultiert die Markov-Bedingung: Eine beliebige Variable  $X_i$  eines Bayes-Netztes ist konditionalisiert auf seine direkten Vorgänger (direkte Ursachen) unabhängig von allen anderen Variablen des Netzwerkes, die nicht Nachfolger (direkte oder indirekte Effekte) von  $X_i$  sind.

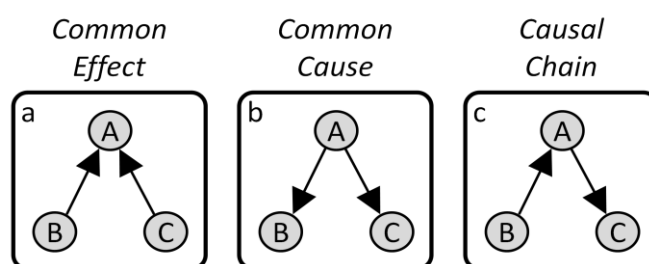


Abbildung 1. Drei grundlegende Kausalstrukturen mit drei Variablen  $A$ ,  $B$  und  $C$ : a) Eine *Common-Effect*-Struktur mit zwei Ursachen, die gemeinsam einen Effekt hervorbringen, b) eine *Common-Cause*-Struktur mit einer Ursache, die zwei Effekte hervorbringt, und c) eine *Causal-Chain*-Struktur, die die Variablen als kausale Kette verbindet.

In Abbildung 1 sind beispielhaft für drei Variablen die grundlegenden Strukturen dargestellt: eine *Common-Effect*-Struktur, eine *Common-Cause*-Struktur und eine *Causal-Chain*-Struktur. In der *Common-Effect*-Struktur sind die beiden Ursachen  $B$  und  $C$  unabhängig voneinander, da sie keine Vorgänger haben; der Effekt  $A$  wird gemeinsam von  $B$  und  $C$  verursacht. In der *Common-Cause*-Struktur verursacht die Ursache  $A$  sowohl den Effekt  $B$  als auch den Effekt  $C$ ; diese sind daher abhängig voneinander (korreliert), konditionalisiert auf  $A$  verschwindet diese Abhängigkeit aufgrund der Markov-Bedingung jedoch. In der *Causal-Chain*-Struktur ist  $B$  die Ursache von  $A$  und  $A$  wiederum die Ursache von  $C$ ; damit sind  $B$

und  $C$  abhängig voneinander. Auch diese Abhängigkeit verschwindet wieder konditionalisiert auf  $A$ , da bei der Betrachtung von  $C$  der Zustand von  $B$  keine Rolle mehr spielt, wenn man den Zustand von  $A$  kennt.

Die Markov-Bedingung erlaubt damit eine einfache Faktorisierung der gemeinsamen Wahrscheinlichkeitsverteilung (*joint probability distribution*) bezüglich der Variablen des Systems. Diese lässt sich dann schreiben als Produkt der bedingten Wahrscheinlichkeiten aller Variablen konditionalisiert auf ihre direkten Vorgänger:

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | \text{parents}(X_i)) \quad (2)$$

Angewendet auf die drei Variablen  $A$ ,  $B$  und  $C$  in Abbildung 1 ergibt sich für die *Common-Effect*-Struktur folglich  $P(A, B, C) = P(A|B, C)P(B)P(C)$ , für die *Common-Cause*-Struktur  $P(A, B, C) = P(A)P(B|A)P(C|A)$  und für die *Causal-Chain*-Struktur  $P(A, B, C) = P(A|B)P(B)P(C|A)$ . Die Markov-Bedingung erlaubt damit insbesondere für große Netzwerke mit vielen Variablen eine kompakte Repräsentation.<sup>12</sup>

Die konkrete Parametrisierung der bedingten Wahrscheinlichkeiten, wie z.B.  $P(A|B)$  oder  $P(A|B, C)$ , ist dabei von den Annahmen abhängig, die über das (Zusammen-)Wirken der Ursachen gemacht werden. In der Psychologie häufig sind z.B. die *Noisy-OR*- oder auch die *Noisy-AND-NOT*-Parametrisierung (siehe u.a. Glymour, 2001 sowie die Ausführungen im folgenden Abschnitt).

---

<sup>12</sup> Es soll nicht unerwähnt bleiben, dass insbesondere die Angemessenheit der Markov-Bedingung auf theoretisch-philosophischer Ebene sehr kontrovers diskutiert wird (siehe v.a. Cartwright, 1983, 1989, 2001, 2002, 2007 sowie Hausman & Woodward, 1999, 2004; Sober, 1987; Woodward, 2003). Danach stellt sich die Frage, ob sich die reale Kausalstruktur der Welt (also z.B. auch physikalische Theorien) angemessen als Bayes-Netz darstellen lässt und welche Bedeutung die dann zu verwendenden Variablen und Relationen haben. Nach Hausman und Woodward (1999; siehe auch Pearl, 2000) ist die Markov-Bedingung lediglich eine computationale Annahme, die jederzeit durch das Einfügen weiterer (unbeobachteter) Variablen hergestellt werden kann. Eine konkrete metaphysische Bedeutung muss diesen nicht zukommen. Dem widerspricht vor allem Cartwright, die eine Theorie von Kausalität, deren Instanzen keine weltlichen Entsprechungen haben, für wertlos hält. Für eine computationale Betrachtung kognitiver Prozesse ist diese Debatte dabei wenig relevant, da eine computationale Theorie kausalen Denkens die kausalen Repräsentationen beschreibt, auf denen das kognitive System operiert. Ob diese eine direkte weltliche Entsprechung haben, ist nicht wichtig, solange sie für den Organismus hilfreich sind, sich in seiner Umwelt zu bewähren.

### 2.2.2 Kausale Bayes-Netze als psychologische Theorie

Schon früh sind Bayes-Netze in ihren qualitativen Eigenschaften und Vorhersagen auf menschliches Kausaldenken und -lernen übertragen bzw. in paralleler Entwicklung aus deren Vorgängern in der Psychologie bekannt geworden (im Wesentlichen auf der Basis der Arbeiten von Pearl, 1988, sowie Spirtes, Glymour und Scheines, 1993, 2000). Insbesondere in Abgrenzung zu assoziationalistischen Lerntheorien wurde die *Causal-Model*-Theorie (Waldmann, 1996, 2000, 2001; Waldmann & Hagmayer, 2001; Waldmann, Hagmayer & Blaisdell, 2006; Waldmann & Holyoak, 1992)<sup>13</sup> entwickelt, die domänenunabhängiges Wissen als Voraussetzung für kausales Lernen hervorhebt, insbesondere das Wissen um die fundamentale Asymmetrie zwischen Ursache und Effekt – eine Unterscheidung, die den assoziationalistischen Lerntheorien fremd ist. So wurden auf der Basis der *Causal-Model*-Theorie und in Widerspruch zu assoziationalistischen Theorien u.a. gezeigt, dass Menschen sensitiv für den Unterschied zwischen prädiktiven Vorhersagen (von der Ursache hin zum Effekt) und diagnostischen Vorhersagen (vom Effekt zur Ursache) sind (siehe u.a. Waldmann, 2000, 2001). Auch konnte z.B. gezeigt werden, dass *cue competition*, ein zentraler Befund aus dem Bereich der klassischen Lerntheorien, mit der Zuschreibung der kausalen Rolle interagiert (Waldmann, 1996, 2000; Waldmann & Holyoak, 1992).

Aber auch im Konkreten sind kausale Bayes-Netze eingesetzt worden, um menschliches Kausallernen zu beschreiben und zu modellieren (siehe u.a. Glymour, 2001, 2003; Gopnik et al., 2004; Griffiths & Tenenbaum, 2005; Steyvers et al., 2003; Waldmann & Martignon, 1998). Im Bereich des rein Daten geleiteten Strukturlernens wird dabei vor allem zwischen den *Constraint-based*-Methoden und den bayesianischen Ansätzen unterschieden.

Die *Constraint-based*-Methoden setzen auf die verschiedenen konditionalen Abhängigkeiten, die die verschiedenen Kausalstrukturen aufgrund der Markov-Bedingung implizieren (sogenannte statistische *Constraints*; siehe Pearl,

---

<sup>13</sup> Für eine ähnliche Übertragung in das Feld der Entscheidungspsychologie in Abgrenzung zu den dort vorherrschenden Theorien, die nicht sensitiv für die Kausalstruktur des Entscheidungsproblems sind, siehe u.a. Hagmayer und Meder (2008), Nichols und Danks (2007) sowie Sloman und Hagmayer (2006).

2000; Spirtes et al., 2000)<sup>14</sup>. Beobachtet man z.B. drei Variablen und ist das Ziel, herauszufinden, wie die drei Variablen kausal verbunden sind, dann sind nach diesem Ansatz die statistischen Abhängigkeiten zu untersuchen. In einer *Common-Effect*-Struktur (siehe Abbildung 1a) sind die beiden Ursachen unabhängig voneinander (unkonditionale Unabhängigkeit), konditionalisiert man diese auf den Effekt, werden sie jedoch abhängig (konditionale Abhängigkeit). In einer *Common-Cause*-Struktur (siehe Abbildung 1b) sind die beiden Effekte abhängig voneinander (unkonditionale Abhängigkeit), da die gemeinsame Ursache eine Scheinkorrelation zwischen diesen erzeugt; konditionalisiert auf die Ursache, werden die Effekte jedoch unabhängig (konditionale Unabhängigkeit), wie sich aus der Markov-Bedingung offenkundig ergibt. Die gleichen Abhängigkeiten ergeben sich aber auch für die *Causal-Chain*-Struktur (siehe Abbildung 1c), die sich damit auf der Basis von rein statistischen Informationen nicht von der *Common-Cause*-Struktur in Abbildung 1b unterscheiden lässt. Solche Strukturen werden als Markov-äquivalent bezeichnet (siehe für die daraus resultierenden Markov-Äquivalenzklassen Abbildung 2).

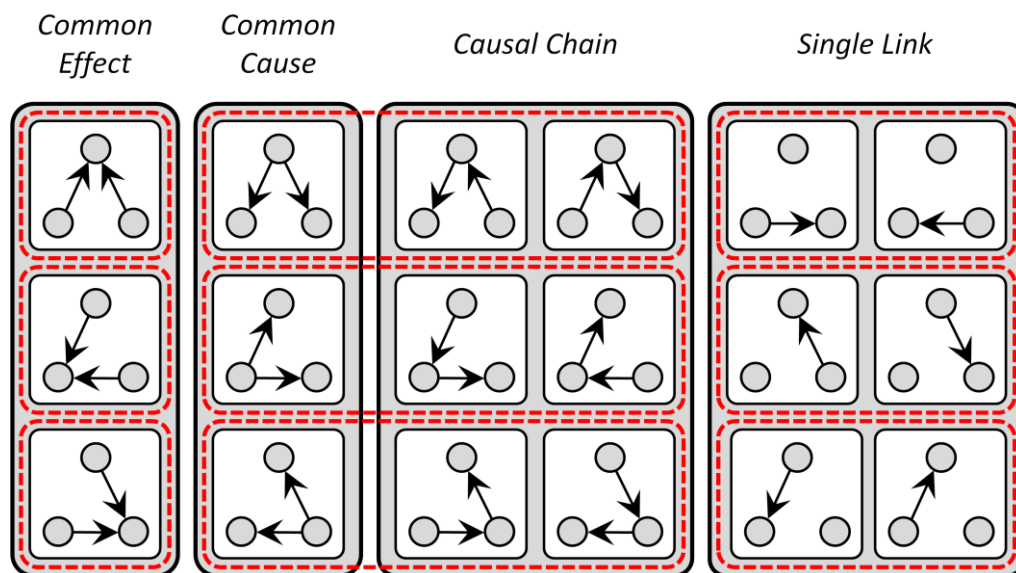


Abbildung 2. Alle Kausalstrukturen über drei Variablen, die einen oder zwei kausale Links enthalten (angelehnt an Steyvers et al., 2003). Die gestrichelten roten Linien kennzeichnen die Markov-Äquivalenzklassen, Strukturen also, die sich auf der Basis rein statistischer Informationen ohne weitere Annahmen nicht voneinander unterscheiden lassen.

<sup>14</sup> Dies setzt jedoch voraus, dass die Abhängigkeiten probabilistisch sind. Für eine Analyse des Problems im Hinblick auf deterministische Systeme siehe Glymour (2007).

---

Abgesehen von diesem Problem leiden *Constraint-based*-Ansätze daran, dass sie eine große Menge an Daten benötigen, um die Abhängigkeiten reliabel zu identifizieren. Nichtsdestotrotz sind sie auch in die Psychologie, insbesondere in die Entwicklungspsychologie, übertragen worden. In sehr einfachen Settings scheinen sogar Kinder sensitiv für die strukturellen Implikationen im Hinblick auf konditionale Abhängigkeiten zu sein (siehe Gopnik et al., 2004; Gopnik, Sobel, Schulz & Glymour, 2001; Kushnir & Gopnik, 2007), auch wenn neuere Befunde (siehe z.B. Lagnado & Sloman, 2002, 2006) es zweifelhaft erscheinen lassen, dass die gefundenen Leistungen wirklich auf eine Sensitivität für konditionale Abhängigkeiten zurückzuführen sind.

Ein in Abgrenzung zu den *Constraint-based*-Methoden entwickelter bayesianischer Ansatz zum Lernen kausaler Strukturen wurde erstmals von Steyvers et al. (2003) empirisch untersucht (für den theoretischen Hintergrund siehe auch Bishop, 2006; Heckerman, 1999; Heckerman, Meek & Cooper, 1999). Diese Ansätze gehen im Grundsatz davon aus, dass das Lernen kausaler Strukturen nichts anderes ist als das Testen verschiedener Strukturhypothesen. Unter Anwendung der Bayes-Formel lässt sich die *A-posteriori*-Wahrscheinlichkeit der möglichen Strukturen gegeben der beobachteten Lerndaten berechnen. Ohne weitere Annahmen lösen aber auch diese Ansätze nicht das Problem der Markov-Äquivalenz, zumal die Anzahl der möglichen Strukturhypothesen (also möglicher Bayes-Netze) exponentiell mit der Anzahl der Variablen wächst. Bereits mit relativ einfachen Strukturen (insbesondere *Common-Cause*- und *Common-Effect*-Strukturen; siehe Abbildung 1), wie sie von Steyvers et al. (2003) genutzt wurden, war die Leistung der Probanden im Hinblick auf die Identifikation der richtigen Kausalstruktur alles andere als überzeugend (ebenda).

Im weiteren Verlauf (und natürlich auch schon vorher) konzentrierte sich die Forschung daher auf die Frage, welche zusätzlichen Informationen (neben Kovariationsdaten) und welches Vorwissen Menschen nutzen, um die kausale Struktur der Welt zu lernen (siehe für einen Überblick Lagnado, Waldmann, Hagmayer & Sloman, 2007) bzw. wie die unüberschaubare Anzahl an potentiell in Frage kommenden Strukturen soweit eingeschränkt werden kann, dass eine Identifikation der richtigen Struktur überhaupt erst möglich ist. Zentrale Bedeu-

tung haben hierbei vor allem die zeitliche Abfolge der Ereignisse, die Möglichkeit, aktiv in das kausale System zu intervenieren, sowie spezifisches und abstraktes Wissen, z.B. in Form kausaler Grammatiken.

*Zeitliche Ereignisabfolge.* Eine wichtige Rolle bei der Entscheidung um die Richtung einer Kausalrelation kommt der zeitlichen Abfolge der Ereignisse zu: Da Ursachen ihre Effekte hervorbringen und nicht umgekehrt, treten Ursachen in der Regel vor Auftreten des Effekts, zumindest aber nicht danach auf. Untersuchungen zeigen, dass Probanden sehr sensitiv für diese zeitliche Asymmetrie sind (Bullock & Gelman, 1979; Lagnado & Sloman, 2006) und dass Informationen über die zeitliche Abfolge von Ereignissen widersprechende Kovariationsinformationen sogar überschreiben können, weil letzteren im Hinblick auf die zugrunde liegende Kausalstruktur bei kleinen Stichproben mehr Unsicherheit anhaftet (Lagnado & Sloman, 2006). Der zeitliche Abstand zwischen Ursache-Ereignis und Effekt-Ereignis spielt auch eine zentrale Rolle, wenn es darum geht, aus einem Strom von Ereignissen die jeweiligen Effekt-Ereignisse ihren Ursachen zuzuordnen, um entsprechende Kontingenzen zu berechnen (Hagmayer & Waldmann, 2002; Krynski, 2006).

*Interventionen.* Neben dem rein passiven Beobachten von Ereignissen und dem darauf aufbauenden Schließen auf zugrunde liegende Kausalstrukturen ist das aktive Eingreifen in Kausalsysteme eine weitere wichtige Möglichkeit, Kausalhypothesen zu testen und kausale Abhängigkeiten zu entdecken (Waldmann & Hagmayer, 2005). Die Konsequenzen von Interventionen sind dabei insbesondere von Pearl (2000; siehe auch Woodward, 2003, 2007) im Bayes-Netz-Framework auf theoretischer Ebene formalisiert worden: Wird in eine Variable aktiv interveniert, ihr Zustand also extern gesetzt, dann verschwindet der Einfluss ihrer Ursachen auf die Variable wie auch auf ihre Effekte. Damit lassen sich durch diese „*graph surgery*“ genannte Operation<sup>15</sup> sogar Markov-äquivalente Strukturen unterscheiden (siehe z.B. Steyvers et al., 2003; Lagnado et al., 2007). Letztendlich sind Interventionen bereits für sich genommen wichtiger Ausdruck des

---

<sup>15</sup> Die Analogie bietet sich an, da formal eine solche Intervention in einem Bayes-Netz resultiert, dem diejenigen Kausalpfeile fehlen, die auf die Variable zeigen, in die interveniert wurde. Es scheint also, als ob diese „wegoperiert“ wurden.

menschlichen Bedürfnisses, aktiv seine Umwelt zu manipulieren. Und so mag es nicht verwundern, dass die Möglichkeit der Intervention das Lernen um die richtige kausale Struktur erheblich beschleunigt oder gar erst ermöglicht (siehe u.a. Campbell, 2007; Hagmayer, Sloman, Lagnado & Waldmann, 2007; Lagnado & Sloman, 2004; Meder, 2006; Meder, Hagmayer & Waldmann, 2008; Schulz, Gopnik & Glymour, 2007; Schulz, Kushnir & Gopnik, 2007; Sloman & Lagnado, 2005; Sommerville, 2007; Waldmann & Hagmayer, 2005). Dabei profitieren nicht nur Menschen von der Möglichkeit des Intervenierens. Auch Ratten scheinen sensitiv für die Konsequenzen ihrer Interventionen in Kausalsysteme zu sein (Blaisdell, Sawa, Leising & Waldmann, 2006; Leising, Wong, Waldmann & Blaisdell, 2008).

*Kausale Grammatiken.* Stehen weder Wissen um die zeitliche Abfolge noch die Möglichkeit einer Intervention und nur eine sehr beschränkte Menge an Beobachtungsdaten zur Verfügung, so scheint ein Lernen zugrunde liegender Kausalrelationen schwierig bis ausgeschlossen. Dennoch identifizieren erwachsene Probanden wie auch Kleinkinder kausale Zusammenhänge in einigen Experimenten bereits nach der Beobachtung weniger Lerntrials (siehe u.a. Meltzoff, 2007; Sobel & Kirkham, 2007; Sobel, Tenenbaum & Gopnik, 2004). Eine Erklärung hierfür scheint zu sein, dass die Probanden abstraktes Wissen über das kausale System nutzen, um den Hypothesenraum einzuschränken oder Beobachtungen restriktiver zu interpretieren. Dieses abstrakte Vorwissen, z.B. dass Risikofaktoren Krankheiten und Krankheiten ihre Symptome verursachen, kann in einer kausalen Grammatik formalisiert werden (siehe Griffiths & Tenenbaum, 2007b; Tenenbaum, Griffiths & Niyogi, 2007; für einen weniger formalisierten, ähnlichen Ansatz siehe bereits Waldmann, 1996, 2007; für verwandte Ansätze siehe auch Kemp, Perfors & Tenenbaum, 2007; Kemp & Tenenbaum, 2008; Kemp & Tenenbaum, in press; Tenenbaum, Griffiths & Kemp, 2006). Eine kausale Grammatik beschreibt dabei in Anlehnung an probabilistische Grammatiken im sprachlichen Bereich (siehe u.a. Chater & Manning, 2006; Xu & Tenenbaum, 2007) einen generativen Algorithmus, der diejenigen Strukturen – hier also die verschiedenen Bayes-Netze bzw. Strukturhypothesen – produziert, über denen die Inferenz stattfindet und der die Strukturen in der Regel auch mit einer *A-priori*-



Wahrscheinlichkeit verbindet (z.B. einfache Strukturen haben eine höhere Wahrscheinlichkeit). Damit wird der Raum der möglichen Hypothesen stark eingeschränkt und eine Inferenz auf der Basis weniger Beobachtungsdaten ermöglicht. Eine solche Grammatik kann z.B. verschiedene abstrakte Klassen von Variablen beschreiben (z.B. Risikofaktoren, Krankheiten, Symptome) und dann die möglichen und Relationen zwischen Variablen dieser Klassen (z.B. ein Risikofaktor kann eine Krankheit verursachen; ein Symptom kann nie eine Krankheit verursachen etc.) sowie mögliche Interaktionen (siehe Griffiths & Tenenbaum, 2007b).

### 2.2.3 Das Causal-Support-Modell (Griffiths & Tenenbaum, 2005)

Einen Brückenschluss zwischen den neueren Bayes-Netz-Modellen zurück zu den ursprünglich in der Psychologie entwickelten Ansätzen ist Griffiths und Tenenbaum (2005) mit ihrem *Causal-Support-Modell* gelungen, indem sie zeigen konnten, dass die psychologischen Maße  $\Delta P$  und *Causal Power* Entsprechungen als Bayes-Netz-Parameter haben. Das *Causal-Support-Modell* formalisiert dabei im Bayes-Netz-Framework Inferenzen bezüglich einer Ursache-Effekt-Relation, wie sie auch Grundlage für die einfachen psychologischen Modelle sind (siehe Abschnitt 2.1), und hebt dabei den Unterschied zwischen *Struktur* – existiert eine Kausalrelation – und *Kausalstärke* – wie stark ist die Relation – hervor. Betrachtet man nun eine (potentielle) Ursache  $C$  und einen (potentiellen) Effekt  $E$  dieser Ursache, dann gibt es auf struktureller Ebene zwei Möglichkeiten: Es besteht keine Kausalrelation zwischen  $C$  und  $E$  (siehe Abbildung 3a), also  $C \nrightarrow E$ , oder es besteht eine Kausalrelation (siehe Abbildung 3b), also  $C \rightarrow E$ . Griffiths und Tenenbaum (2005) zeigen, dass menschliche Kausalurteile stärker mit der Strukturfrage, also ob eine Kausalrelation besteht, korrespondieren als mit der Frage nach der Stärke einer kausalen Verbindung. Die Strukturfrage formalisieren sie dabei über das Log-Likelihood-Verhältnis der Beobachtungsdaten  $D$  gegeben einer Kausalrelation und gegeben keiner Relation:

$$support = \log \frac{P(D|C \rightarrow E)}{P(D|C \nrightarrow E)} \quad (3)$$

Dieses *Causal Support* genannte Maß ist umso größer, je wahrscheinlicher eine Kausalrelation zwischen  $C$  und  $E$  gegeben der beobachteten Daten ist.

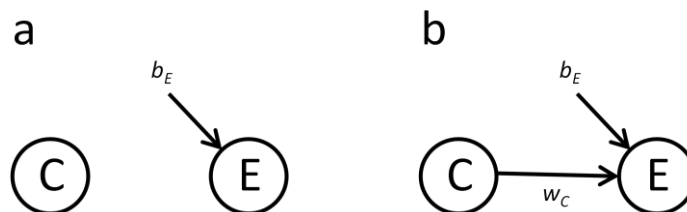


Abbildung 3. Die beiden möglichen Grundstrukturen bei Betrachtung einer Ursache  $C$  und eines Effekts  $E$ : a) Zwischen der Ursache  $C$  und dem Effekt  $E$  besteht keine Kausalrelation. Der Effekt wird nur vom Hintergrund mit der Wahrscheinlichkeit  $b_E$  (Basisrate) produziert. b) Es besteht eine Kausalrelation zwischen  $C$  und  $E$ . Der Effekt wird damit sowohl von der Ursache mit Stärke  $w_C$  als auch vom Hintergrund erzeugt. Weitere Erläuterungen siehe Text.

Viel interessanter als diese Erkenntnis ist dabei die Art, wie Griffiths und Tenenbaum (2005) das zugrunde liegende Likelihood formalisieren. Betrachtet man dazu die Struktur mit Kausalrelation zwischen  $C$  und  $E$  (Abbildung 3b), dann stellt sich die Frage, wie die Ursache  $C$  mit dem Hintergrund, der hier als Basisrate  $b_E$  dargestellt ist<sup>16</sup>, interagiert. Sei  $w_C$  dazu die Wahrscheinlichkeit mit der die Ursache  $C$  den Effekt  $E$  verursacht. Unter der Annahme, dass sowohl Hintergrund als auch  $C$  den Effekt *unabhängig* produzieren, ergibt sich für die Wahrscheinlichkeit der Anwesenheit des Effekts gegeben der Anwesenheit von  $C$  nach den Regeln der Wahrscheinlichkeitsrechnung für die Disjunktion zweier unabhängiger Ereignisse ( $E$  wird von  $C$  produziert mit Wahrscheinlichkeit  $w_C$  sowie  $E$  wird vom Hintergrund erzeugt mit Wahrscheinlichkeit  $b_E$ ):  $P(E^+|C^+) = b_E + w_C - b_E w_C$  und entsprechend allgemeiner:  $P(E^+|C) = 1 - (1 - b_E)(1 - w_C)^C$  (diese Verknüpfungsart wird auch *Noisy-OR* genannt und liegt auch der *Power-PC*-Theorie zugrunde, siehe Glymour, 1998, 2001). Griffiths und Tenenbaum konnten zeigen, dass der Maximum-Likelihood-Schätzer für die Kausalstärke  $w_C$  unter der *Noisy-OR*-Annahme  $p_C$ , also *Causal Power* (siehe Abschnitt 2.1.2), entspricht. Nimmt man für die Verknüpfung von  $C$  und Hintergrund eine lineare Funktion an, also  $P(E^+|C^+) = b_E + w_C$ , dann ist der Maximum-Likelihood-Schätzer für  $w_C$  entsprechend  $\Delta P$  (siehe Abschnitt 2.1.2). Damit demonstrieren Griffiths und

<sup>16</sup> Es handelt sich um eine Vereinfachung unter der Annahme, dass die Hintergrundursache immer anwesend ist und mit der Wahrscheinlichkeit  $b_E$  den Effekt hervorbringt. Damit ist  $b_E$  die Wahrscheinlichkeit, mit der der Effekt ohne Anwesenheit der Ursache auftritt. Sie wird daher auch Basisrate genannt.

Tenenbaum, dass sich die beiden wichtigsten psychologischen Modelle elementaren Kausallernens als zwei verschiedene Interpretationen von Kausalstärke unter unterschiedlichen funktionalen Verknüpfungsannahmen im Bayes-Netz-Framework vereinigen lassen.

Des Weiteren haben Griffiths und Tenenbaum gezeigt, wie sich Vorwissen um die kausalen Parameter in die Inferenz einfügen lässt: Da weder  $w_C$  noch  $b_E$  bekannt sind, müssen diese bei der Bestimmung des Likelihoods geschätzt werden. Dazu werden diese als unbekannte Variablen eingeführt und ausintegriert<sup>17</sup>

$$P(D|C \rightarrow E) = \iint P(D|C \rightarrow E, w_C, b_E)P(w_C)P(b_E)dw_Cdb_E \quad (4)$$

Die Wahrscheinlichkeitsverteilungen  $P(w_C)$  und  $P(b_E)$  beschreiben dabei die Vorannahmen bezüglich der Ausprägung dieser Parameter (*A-priori*-Verteilungen), sie werden in der Regel als Beta-Verteilungen<sup>18</sup> dargestellt, z.B.:  $w_C \sim \text{Beta}(\alpha_{w_C}, \beta_{w_C})$ . Die beiden Hyperparameter  $\alpha_{w_C}$  und  $\beta_{w_C}$  stellen damit das abstrakte Vorwissen über den Parameter  $w_C$  dar. Griffiths und Tenenbaum (2005) nutzen für die Modellierung ihrer Daten aus Vereinfachungsgründen uninformative Beta-Verteilungen, indem sie beide Hyperparameter auf 1 gesetzt haben. Eine solche Beta-Verteilung ordnet jeder Ausprägung des zugrunde liegenden Parameters die gleiche Wahrscheinlichkeit zu, um auszudrücken, dass das kognitive System keinerlei *A-priori*-Präferenz für bestimmte Ausprägungen hat. Im Grundsatz kann damit jedoch jedwedes Vorwissen implementiert werden. So haben zum Beispiel Lu et al. (2008) aufbauend auf dem *Causal-Support*-Modell gezeigt, dass menschliche Kausalurteile sich viel besser abbilden lassen unter der Annahme, dass Ursachen selten und stark sind (sogenannter *SS-Prior* [*strong and sparse*]), also eine starke Präferenz für entweder ein hohes  $w_C$  oder ein hohes  $b_E$ .

<sup>17</sup> Dieses Vorgehen entspricht dabei der doppelten Anwendung der allgemeinen Regel  $P(X) = \int P(X|\theta)P(\theta)d\theta$ ; siehe dazu allgemein auch Gelman, Carlin, Stern und Rubin (2004) sowie speziell im Bereich der kognitiven Modelle Chater, Tenenbaum und Yuille (2006), Griffiths, Kemp und Tenenbaum (2008) sowie Griffiths und Yuille (2008).

<sup>18</sup> Eine Beta-Verteilung ordnet einer Variablen  $x$  im Definitionsbereich 0 bis 1 eine Wahrscheinlichkeitsdichte zu:  $f(x; \alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}$  mit  $B(\alpha, \beta) = \int x^{\alpha-1} (1-x)^{\beta-1} dx$ . Die beiden Parameter  $\alpha$  und  $\beta$  beschreiben die Form der Verteilung. Die Beta-Verteilung ist konjugierte Verteilung zur Binomial-Verteilung, was das Rechnen mit ihr erheblich vereinfacht.

### 2.3 Empirische Untersuchungen zur Markov-Bedingung

Gleichwohl die Markov-Bedingung eine zentrale Rolle in der Bayes-Netz-Theorie spielt und sie Voraussetzung für die oben dargestellten Modelle ist, wurde in den meisten Arbeiten zum Thema ihre psychologische Validität schlichtweg angenommen, aber nie selbst untersucht. Die Frage ist also, ob Menschen die in der Markov-Bedingung beschriebene konditionale Unabhängigkeit der Ereignisse auch repräsentieren. Forschungen in diesem Bereich sind dabei erst in den letzten Jahren, insbesondere in der Arbeitsgruppe um Bob Rehder, aufgenommen worden, die im Folgenden dargestellt werden sollen.

Um die Intuitionen der Versuchspersonen hinsichtlich der Markov-Bedingung zu untersuchen, entwickelten Rehder und Burnett (2005) eine kausale Inferenzaufgabe. Da die Studien an der Schnittstelle zwischen Kausaldenken und Kategorisierung angesiedelt waren, wurden die Probanden in einem ersten Experiment in Szenarios mit einer bestimmten Kategorie eingeführt (z.B. „*Lake Victoria Shrimp*“, „*Kehoe Ants*“ oder „*Neptune Computers*“). Jedes Exemplar einer solchen Kategorie hatte vier Merkmale (z.B. „*ACh-level: high vs. low*“, „*Flight response: short- vs. long-lasting*“, „*Sleep cycle: accelerated vs. decelerated*“ und „*Body weight: high vs. low*“ für die Kategorie „*Lake Victoria Shrimp*“), wobei jeweils eine der Ausprägungen als die typische Ausprägung beschrieben wurde, die in 75% der Fälle bei einem Exemplar der Kategorie vorliegt. Die Merkmale waren des Weiteren kausal als *Common-Cause*-Struktur<sup>19</sup> mit einer gemeinsamen Ursache und drei Effekten instruiert (siehe Abbildung 4), wobei die kausalen Relationen verbal beschrieben wurden (z.B. „Eine große Menge des ACh-Neurotransmitters verursacht eine lang andauernde Fluchtreaktion. Das elektrische Signal an die Muskeln dauert aufgrund des Überschusses an Neurotransmittern länger.“; siehe Anhang A in Rehder & Burnett, 2005; Übersetzung des Autors). Die typische Ausprägung des Ursache-Merkmals hat dabei immer die typische Ausprägung des jeweiligen Effekt-Merkmals verursacht.

---

<sup>19</sup> Rehder und Burnett (2005) haben auch weitere Strukturen getestet, die aus Vereinfachungsgründen aber nicht dargestellt werden.

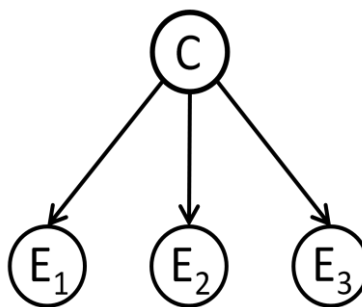


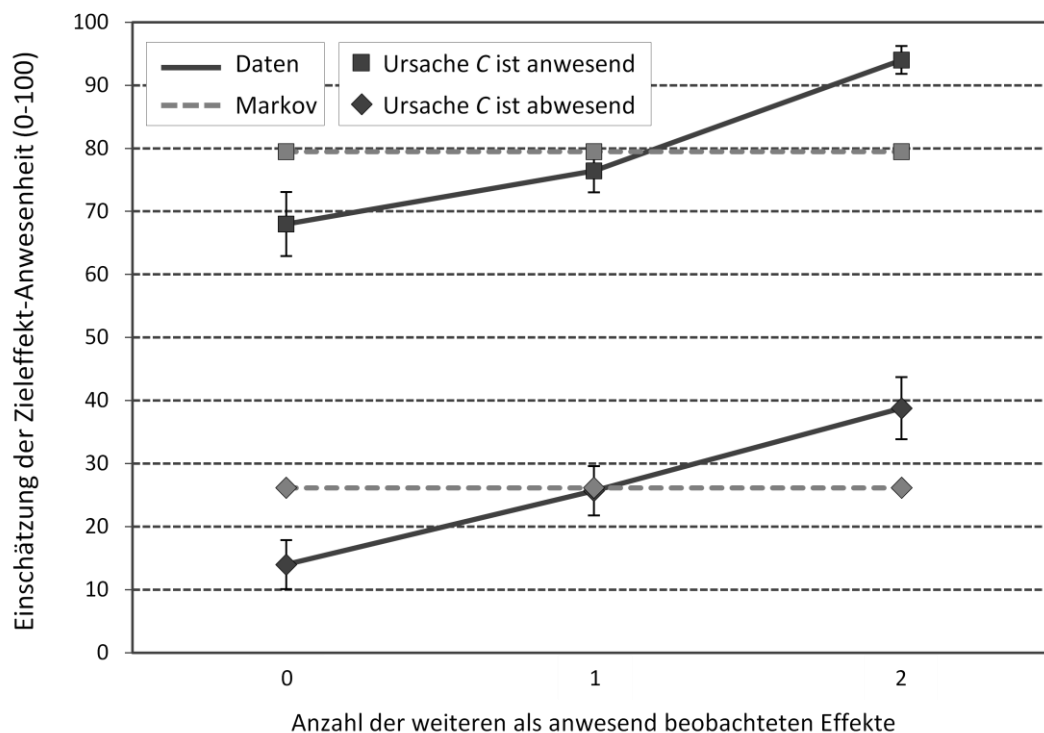
Abbildung 4. Einfache *Common-Cause*-Struktur mit einer Ursache  $C$  und drei Effekten,  $E_1$ ,  $E_2$  und  $E_3$ , wie sie von Rehder und Burnett (2005) in den beschriebenen Experimenten verwendet wurde.

In der Testphase<sup>20</sup> wurden den Probanden Testfälle vorgelegt, die den Zustand des Ursache-Merkmals beschrieben (anwesend = typisch bzw. abwesend = untypisch) sowie die Zustände zweier Effekt-Merkmale. Aufgabe der Probanden war es, auf einer Skala von 0 bis 100 einzuschätzen, wie wahrscheinlich es sei, dass das dritte Effekt-Merkmal, dessen Status nicht bekannt war, anwesend ist (dieses also die typische Ausprägung hat). Der Markov-Bedingung entsprechend dürften diese Urteile nur vom Zustand der Ursache, aber nicht vom Wissen um den Zustand der weiteren Effekte der *Common-Cause*-Struktur abhängig sein, da der Zieffekt konditionalisiert auf die Ursache unabhängig von den weiteren Variablen des Systems ist.

In Abbildung 5 sind die Ergebnisse des Experiments (durchgezogenen Linien) sowie eine Vorhersage auf der Basis der Markov-Bedingung (gestrichelte Linien) abgetragen. Die oberen beiden Linien stellen dabei die Einschätzungen für den Fall der Anwesenheit der Ursache und die unteren beiden Linien entsprechend den Fall ihrer Abwesenheit dar. Auf der X-Achse abgetragen findet sich die Anzahl der weiteren als anwesend beobachteten Effekte, im vorliegenden Fall also null, eins oder zwei (da der dritte Effekt unbeobachtet ist, können also höchstens zwei weitere Effekte anwesend sein). Da der Markov-Bedingung nach die Inferenz über den Zieffekt unabhängig von den anderen Effekten sein sollte, deren Anwesenheit oder Abwesenheit mithin also keinen Einfluss haben dürfte, entspricht die Markov-konforme Vorhersage zwei horizontalen Linien. Im Fall,

<sup>20</sup> Rehder und Burnett haben noch mehr als das hier Beschriebene abgefragt, so zum Beispiel die Wahrscheinlichkeit der Zugehörigkeit zur Kategorie etc. Die weiteren Fragen sind aber im Hinblick auf die Markov-Bedingung nicht interessant. Auf deren Darstellung wird daher aus Gründen der Übersichtlichkeit verzichtet.

dass die Ursache anwesend ist, sollten die Beurteilungen konstant hoch ausfallen; im Fall, dass diese abwesend ist, entsprechend konstant niedrig (auf dem Niveau der Basisrate des Zieleffekts). Empirisch jedoch zeigen sich sehr deutliche Abhängigkeiten: Ausgehend von null als anwesend beobachteten weiteren Effekten steigen die Einschätzungen bezüglich der Anwesenheit des Zieleffekts mit deren steigender Anzahl für beide Zustände der Ursache stark an.



*Abbildung 5.* Die Ergebnisse der *Common-Cause*-Bedingung von Experiment 1 aus Rehder und Burnett (2005) für natürliche Kategorien. Den Daten (durchgezogene Linien) gegenübergestellt ist die Vorhersage eines Bayes-Netzes, also auf der Basis der Markov-Bedingung (gestrichelte Linie), wie sie einfach aus der Mittelung der Datenpunkte je Zustand der Ursache gewonnen wurden. Eingeschätzt werden sollte die Wahrscheinlichkeit der Anwesenheit eines unbeobachteten Zieleffekts (Skala von 0 bis 100) gegeben der Anwesenheit oder Abwesenheit der Ursache *C* (die beiden oberen vs. die beiden unteren Linien) sowie der Anzahl der weiteren als anwesend beobachteten Effekte (*X*-Achse: 0, 1 oder 2). Die Fehlerbalken sind Standardfehler des Mittelwerts; die Daten sind Rehder und Burnett (2005) entnommen.

Auch in einem weiteren Experiment, in dem Rehder und Burnett (2005) arbiträre Kategorien mit abstrakten Merkmalen (*A*, *B*, *C* und *D*) benutzen und auch die kausalen Verbindungen abstrakt beschrieben wurden (*A* verursacht *B*, *C* und *D*), fanden sich die gleichen Abhängigkeiten in ähnlichem Ausmaß. Damit konnte

ausgeschlossen werden, dass es sich lediglich um ein Artefakt dahingehend handelt, dass die Versuchspersonen in das natürliche Material, welches im ersten Experiment verwendet wurde, auf der Basis ihres Hintergrundwissens, z.B. über die Wirkweise von Neurotransmittern, zusätzliche Relationen zwischen den Merkmalen angenommen haben, die verantwortlich für die Abhängigkeiten sind.

Als Erklärung haben Rehder und Burnett (2005) zwei verschiedene Modelle vorgeschlagen, die die zugrunde liegende *Common-Cause*-Struktur computational erweitern. Es handelt sich zum einen um das *Underlying-Mechanism*-Modell (siehe Abbildung 6a) und zum anderen um das *Feature-Uncertainty*-Modell (siehe Abbildung 6b).

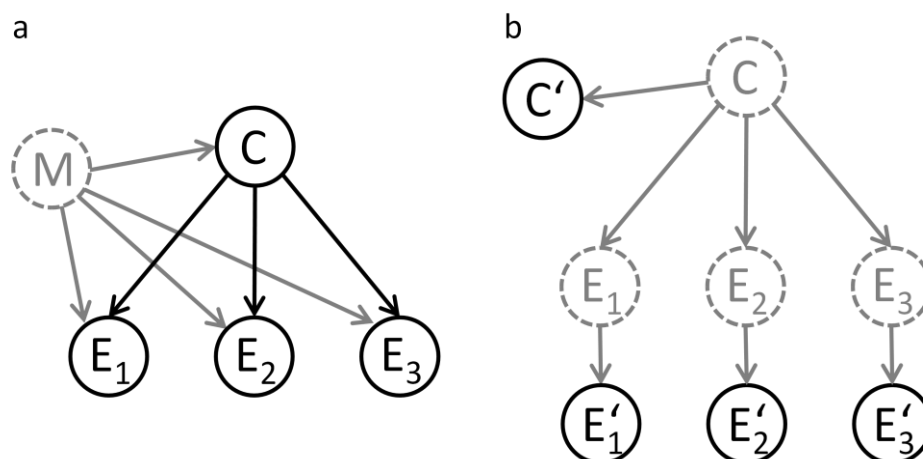


Abbildung 6. Die von Rehder und Burnett (2005) vorgeschlagenen Modelle zur Erklärung von Markov-Verletzungen (in Anlehnung an Abbildung 8, S. 284, ebenda). Dies sind (a) das *Underlying-Mechanism*-Modell, in welchem die Ausprägung der vier Merkmale als von einem zugrunde liegenden kategorialen Mechanismus  $M$  erzeugt gedacht werden, und (b) das *Feature-Uncertainty*-Modell, nach welchem die vier kausalen Variablen nicht direkt beobachtet sind und mit deren Beobachtung Unsicherheit über die wahre Ausprägung verbunden ist. Weitere Erläuterung siehe Text.

Das *Underlying-Mechanism*-Modell geht von der Annahme aus, dass die Exemplare der Kategorie von einem der Kategorie zugrunde liegenden generativen Mechanismus produziert werden, der meistens eher typische und manchmal aber auch atypische Exemplare hervorbringt. Beobachtet man z.B. bei Anwesenheit der Ursache  $C$  die Abwesenheit zweier Effekte, so handelt es sich eher um ein atypisches Exemplar, der unbeobachtete Effekt sollte daher auch eher abwe-

send sein. Die reine (Markov-konforme) Kausalinferenz interagiert hier also mit der Inferenz bezüglich der Kategorie, welche Markov-Verletzungen auf der Ebene der beobachtbaren Variablen erzeugt.

Im *Feature-Uncertainty*-Modell werden die Variablen des kausalen Systems, also  $C$ ,  $E_1$ ,  $E_2$  und  $E_3$ , als nicht direkt beobachtbar gedacht. Beobachtet werden stattdessen die beobachtbaren Instanzen  $C'$ ,  $E'_1$ ,  $E'_2$  und  $E'_3$ , die aber mit einer gewissen Unsicherheit hinsichtlich der wahren Ausprägung der zugrunde liegenden Variablen verbunden sind. Beobachtet man hier wieder z.B. bei Anwesenheit der Ursache-Instanz  $C'$  die Abwesenheit zweier Effekt-Instanzen, so kann man dies darauf zurückführen, dass die wahre Ursache  $C$  möglicherweise gar nicht anwesend ist und der nicht beobachtete Effekt daher auch nicht anwesend sein sollte. Die (wieder Markov-konforme) Kausalinferenz interagiert in diesem Modell mit der Unsicherheit hinsichtlich der Ausprägung der beobachtbaren Variablen.

Beide Modelle machen für die beiden durchgeführten Experimente die gleichen Vorhersagen. In weiteren Experimenten testeten Rehder und Burnett (2005) die Vorhersagen dieser Modelle in neuen Kontexten und kamen zu dem Schluss, dass das *Underlying-Mechanism*-Modell die Befunde am besten erklärt.

#### **2.4 Kausale Prozesse, Fehlerstrukturen und die Markov-Bedingung**

Die Ergebnisse von Rehder und Burnett (2005) zeigen deutlich, dass die *A-priori*-Annahme der Validität der Markov-Bedingung unangemessen ist. Walsh und Sloman (2004) sowie Rehder (2006) konnten des Weiteren zeigen, dass Markov-Verletzungen auch außerhalb des kategorialen Kontextes auftreten (z.B. bei der Beurteilung der konditionalen Abhängigkeit ökonomischer Variablen), ohne jedoch eine Erklärung hierfür anzubieten (außer, dass menschliches Kausaldenken möglicherweise schlicht irrational ist). Das von Rehder und Burnett (2005) entwickelte Modell lässt sich seiner Idee nach aber nur auf kausal verknüpfte Merkmale von Kategorien anwenden und ist damit nicht geeignet, Markov-Verletzungen als allgemeines Phänomen zu beschreiben. Die zugrunde liegende Erklärung ist auch wie oben beschrieben nicht direkt kausaler Natur: Im *Underlying-Mechanism*-Modell wie auch im *Feature-Uncertainty*-Modell interagiert die (immer noch) Markov-konforme Kausalinferenz mit anderen Inferenzprozessen, wie



z.B. bezüglich der Typikalität eines Exemplars oder bezüglich der Unsicherheit über die Ausprägung der beobachteten Variablen.

Das Phänomen der Markov-Verletzung könnte jedoch auch in der Kausalinferenz selbst liegen. Dies würde erklären, warum sie nicht dem *Underlying-Mechanism-Modell* entsprechend nur auf kategoriale Kontexte beschränkt ist. Betrachten wir dazu die Relation zwischen einer Ursache  $C$ , z.B. einem Virus, und einem Effekt  $E$ , z.B. einem Symptom wie Fieber. Wenn die Ursache  $C$  ihren Effekt nunmehr in 80% der Fälle ihrer Anwesenheit hervorbringt (dies entspricht dann einer Kausalstärke von 0.8), dann heißt dies auch, dass sie diesen in 20% der Fälle nicht hervorbringt. Warum sie in diesen Fällen versagt, ist aber nicht Bestandteil der bis dato dominierenden kovariationsbasierten Ansätze sowie im Prinzip auch nicht der Bayes-Netz-Theorien. Es ist vielmehr Frage des dieser rein statistischen Kausalrelation zugrunde liegenden Mechanismus (siehe dazu auch Abschnitt 2.1.3), also wie die Ursache ihren Effekt hervorbringt und warum sie dies in einigen Fällen nicht schafft. Dabei könnte man sich natürlich auch rein probabilistische Mechanismen, wie zum Beispiel in der Quantenmechanik, vorstellen<sup>21</sup>. Ein Versagen wäre demnach Zufall und hätte keinen höheren Grund. Forschung zeigt jedoch, dass Menschen ein eher quasi-deterministisches Weltverständnis haben (Griffiths & Tenenbaum, 2007; Richardson, Schulz & Gopnik, 2007; Schulz & Sommerville, 2006; siehe auch Luhmann & Ahn, 2005; für eine Formalisierung in funktionalen Kausalmodellen siehe Pearl, 2000). Das heißt, dass ein probabilistischer Zusammenhang zwischen Ursache  $C$  und Effekt  $E$  auf nicht beobachtete präventive Ursachen zurückzuführen ist, auf Ursachen also, die  $C$  daran hindern, den Effekt  $E$  hervorzubringen. Das Komplement zur Kausalstärke ist damit Ausdruck der Häufigkeit/Stärke solcher präventiven Ursachen (für eine formalere Darstellung siehe Abschnitt 4.1).

Betrachtet man diesbezüglich nun eine *Common-Cause*-Struktur, wie sie von Rehder und Burnett (2005) verwendet wurde, z.B. einen Virus als Ursache dreier Symptome wie Fieber, Kopfschmerzen und Durchfall, dann erscheint es plausibel, dass die Frage des zugrunde liegenden Mechanismus und damit der Frage möglicher präventiver Ursachen einen Einfluss auf die Inferenzen in der

---

<sup>21</sup> Der *Power-PC*-Theorie von Cheng (1997) liegt ein solch probabilistisches Kausalitätsverständnis zugrunde (siehe für eine Kritik u.a. Luhmann & Ahn, 2005).

---

Gesamtstruktur hat. Die Markov-Bedingung verlangt, dass diese potentiellen Fehler für die einzelnen Links Virus–Fieber, Virus–Kopfschmerzen und Virus–Durchfall unabhängig sind, die gedachten präventiven Ursachen also unabhängig voneinander wirken. Im genannten Beispiel könnte man eine diesbezügliche Intuition haben: Der Virus zirkuliert im Blut und greift das Lymphsystem an (führt zu Fieber), schädigt die Blut-Hirn-Schranke (führt zu Kopfschmerzen) und schwächt den Magen-Darm-Trakt (führt zu Durchfall). Dass der Virus ein Symptom nicht immer hervorbringt, kann in diesem Fall an spezifischen, unabhängigen präventiven Ursachen an den verschiedenen Wirkorten liegen. In diesem Fall sollten die Inferenzen hinsichtlich der Anwesenheit eines Effekts gar nicht (oder zumindest nur wenig) von der Anwesenheit oder Abwesenheit der anderen Effekte abhängig sein, da deren Zustand nur etwas über die für den jeweiligen Effekt spezifische präventive Ursache aussagt, nicht aber über diejenige, die zum Zieffekt gehört.

Das Beispiel lässt sich aber auch mit anderen Intuitionen über die zugrunde liegenden Kausalmechanismen verbinden: Zum Beispiel, dass der Virus in Leberzellen eindringt und dort Gene aktiviert, die die betroffenen Zellen eine Substanz produzieren lassen, die die entsprechenden Symptome hervorbringt. Eine mögliche präventive Ursache wären Anti-Körper, die den Virus davon abhalten, in die Leberzellen einzudringen und damit die schädigende Substanz zu produzieren. In diesem Fall würde die Abwesenheit der Symptome für die Anwesenheit der präventiven Ursache sprechen, die sich auch auf den Zieffekt auswirkt. Auf der Ebene der *beobachtbaren* Variablen wären die Inferenzen bezüglich der Effekte nicht mehr konditional unabhängig.

Der Unterschied zwischen den beiden Beispielen liegt dabei offenkundig in der Frage, ob die potentiellen Fehlerquellen eher effektseitig, was konditionale Unabhängigkeit impliziert, oder ursacheseitig lokalisiert sind, was zu konditionalen Abhängigkeiten führt. Dem Beispiel folgend, mag der Schlüssel zur Erklärung von Markov-Verletzungen also in den kausalen Prozessen liegen: *Wie* erzeugt die Ursache ihre Effekte und welche Probleme können hierbei auftreten. – Dabei muss es sich nicht um konkretes Wissen handeln. Allein das Wissen, dass es solche, auf alle Effekte wirkenden präventiven Ursachen geben kann, würde dann

auch in sehr abstrakten Settings, wie zum Beispiel in dem zweiten Experiment von Rehder und Burnett (2005) zu konditionalen Abhängigkeiten führen (die Markov-Bedingung wäre damit gerade keine Default-Annahme im menschlichen Kausaldenken).

Ist diese eben entwickelte Idee richtig, dann müssten Manipulationen bezüglich der involvierten kausalen Prozesse auch in entsprechend verändertem Bewertungsverhalten resultieren, wie es im nächsten Kapitel untersucht werden soll: Legen die Prozesse eine effektseitige Fehlerattribution nahe, dann sollten die Bewertungen der Anwesenheit eines Zieleffekts deutlich weniger vom Zustand weiterer Effekte einer gemeinsamen Ursache abhängig sein, als wenn die mit dem System assoziierten Prozesse eher eine ursacheseitige Fehlerattribution nahe legen. Diese Überlegungen und diese Vorhersage heben sich damit insoweit von den existierenden Untersuchungen von Rehder und Burnett (2005) ab, also dass diese zwar ebenso über verschiedene Cover-Stories verschiedene Kausalprozesse instruiert, diese aber ausbalanciert und nicht systematisch im Hinblick auf deren Wirkungen auf die Beurteilungen manipuliert haben. Zwar ist die Idee, dass sich Annahmen über die kausalen Prozesse systematisch auf das Lernen von Kausalrelationen auswirken, auch bereits in den mechanismusbasierten Ansätzen (siehe Abschnitt 2.1.3) vorhanden. Der dortige Fokus liegt aber auf der Frage (in Anlehnung an die kovariationsbasierten Ansätze), wie Probanden einzelne Ursache-Effekt-Relationen lernen, nicht jedoch darauf, wie die Annahmen über Mechanismen Inferenzen in komplexeren Kausalstrukturen, wie z.B. in einer *Common-Cause*-Struktur, beeinflussen.

## 2.5 Zusammenfassung

Im vorliegenden Kapitel wurden die theoretischen und empirischen Hintergründe für die Beschäftigung mit der Markov-Bedingung als Teil einer psychologischen Theorie kausalen Denkens zusammengefasst. Als Ausgangspunkt wurde die Literatur zum kausalen Lernen und Denken dargestellt und ein Überblick über die wichtigsten psychologischen Theorien, wie die assoziationstheoretischen Ansätze, die kovariationsbasierten Ansätze, wie die  $\Delta P$ -Regel und *Causal Power*, sowie die mechanismusbasierten Ansätze gegeben. Diese widmen sich vor allem der Frage, wie Probanden einzelne Ursache-Effekt-Relationen lernen.

---

Daran anschließend wurde die Bayes-Netz-Theorie eingeführt, die sich mathematisch als Verknüpfung von gerichteten kausalen Graphen und dem Wahrscheinlichkeitskalkül darstellt. Zentrale Annahme ist die Markov-Bedingung, nach der eine Variable in einer Kausalstruktur konditionalisiert auf ihre direkten Ursachen unabhängig von den anderen Variablen ist, die nicht Nachfolger der Variablen sind. Darauf aufbauend wurde gezeigt, wie Kausalstrukturen mit den *Constraint-based*-Methoden und mittels bayesianischen Hypothesentesten entdeckt werden können und dass Probanden in solchen Aufgaben oft sehr schlecht abschneiden. Menschen scheinen daher auch weitere Informationen, wie die zeitliche Abfolge von Ereignissen und das Ergebnis von Interventionen, und abstraktes Vorwissen, wie es in kausalen Grammatiken formal beschrieben werden kann, zu nutzen, um Kausalrelationen zu identifizieren.

Als Brückenschluss zurück zu den psychologischen Ansätzen wurde das *Causal-Support*-Modell von Griffiths und Tenenbaum (2005) eingeführt. Dieses formalisiert das Lernen elementarer Ursache-Effekt-Relationen im Bayes-Netz-Framework. Griffiths und Tenenbaum konnten zeigen, dass die psychologischen Maße  $\Delta P$  und *Causal Power* beide Maximum-Likelihood-Schätzer der Kausalstärke unter unterschiedlichen Annahmen bezüglich der funktionalen Verknüpfung von Ursache und Hintergrund sind.

Des Weiteren sind die empirischen Arbeiten von Rehder und Burnett (2005) genauer vorgestellt worden. Diese nähren Zweifel an der Markov-Bedingung als Teil einer psychologischen Theorie des kausalen Denkens. Die Beurteilung der Anwesenheit eines Effektes in einer *Common-Cause*-Struktur gegeben der An- oder Abwesenheit seiner Ursache hängt im Widerspruch zur Markov-Bedingung sehr wohl und gleichsam auch sehr deutlich vom Status der weiteren Effekte der gemeinsamen Ursache ab. Darauf aufbauend und in Abgrenzung zu den akasalen Erklärungsversuchen von Rehder und Burnett wurde die Idee entwickelt, dass Probanden bei Kausalinferenzen systematisch weiteres Wissen, insbesondere um die zugrunde liegenden kausalen Prozesse und die damit verbundene Fehlerstruktur, miteinbeziehen. Dies soll im folgenden Kapitel genauer untersucht werden.

### 3 Basisexperimente

Wie im vorhergehenden Kapitel ausgeführt, stellt sich die Frage, inwieweit die Urteile der Versuchspersonen über unbeobachtete Ereignisse in einer kausalen Domäne von deren Annahmen über die zugrunde liegenden kausalen Prozesse abhängig sind. In Abschnitt 2.4 ist die allgemeine Vorhersage abgeleitet worden, dass, wenn diese Prozesse eine effektseitige Fehlerattribution nahelegen, die Bewertungen hinsichtlich der Anwesenheit eines unbeobachteten Zieleffekts deutlich weniger vom Zustand weiterer Effekte einer gemeinsamen Ursache abhängig sein sollten, als wenn die zugrunde liegenden Mechanismen eher eine ursacheseitige Fehlerattribution nahelegen.

Zur Prüfung dieser Hypothese haben wir das experimentelle Paradigma der Gedanken lesenden Außerirdischen („*Alien-Mind-Reader-Task*“; Steyvers et al., 2003) als kausale Domäne gewählt, da hier die Einflüsse von Vorwissen minimal sind und sich das Wissen der Versuchspersonen um die zugrunde liegenden kausalen Mechanismen gut manipulieren lässt. In ihren Experimenten haben Steyvers und Kollegen die Versuchspersonen mit drei Außerirdischen konfrontiert. Jedes Alien dachte in jedem Trial an ein 3-buchstabiges sinnfreies Wort aus einem begrenzten Vokabular (z.B. „POR“, „TUS“, ...). Außerdem wurde den Versuchspersonen erklärt, dass einige Aliens in der Lage seien, die Gedanken von anderen Außerirdischen zu lesen: Wenn Alien A die Gedanken von Alien B lesen kann, dann denkt A normalerweise das gleiche Wort wie Alien B. Des Weiteren wurde den Probanden mitgeteilt, dass das Gedankenlesen manchmal misslingt, wenn das Gedanken lesende Alien gerade un aufmerksam ist. Nachdem die Versuchspersonen einige Kommunikationsmuster gesehen hatten, sollten sie angeben, welches/welche der Aliens Gedanken lesen können.

Diese Aufgabe mag auf den ersten Blick zwar nicht aussehen wie eine kausale Aufgabe, doch ist sie äquivalent zur Identifikation von Kausalstrukturen: Die drei Aliens spannen ein dreiknotiges Netzwerk auf und die Identifikation von Gedankenlesern ist gleichbedeutend mit der Identifikation von kausalen Links in diesem Netzwerk. Da die Gedanken eines Gedanken lesenden Aliens statistisch abhängig sind von den Gedanken desjenigen Aliens, dessen Gedanken gelesen werden, kann diese Relation als kausaler Link interpretiert werden, der zum Ge-

---

danken lesenden Alien zeigt. Das heißt, das Gedanken lesende Alien ist der Effekt und das Alien, von dem gelesen wird, die Ursache. Das erscheint zwar kontraintuitiv, weil schließlich das Gedanken lesende Alien aktiv ist („es liest“) und dieses daher etwas zu verursachen scheint und nicht andersherum, doch auf rein statistischer Betrachtungsebene der Ereignisse – der Gedanken der Aliens – und aus der Perspektive von kausalen Bayes-Netzen ist diese tiefere kausale Ebene nicht relevant. Die Begriffe „Ursache“ und „Effekt“ sind rein statistisch auf der Basis konditionaler Abhängigkeiten definiert. – Genau dies ist eines der Probleme dieses Ansatzes, wenn man ihn auch auf der inhaltlichen Ebene als psychologische Theorie des kausalen Denkens einsetzen möchte.

Die „Domäne“ Gedanken lesender Aliens hat – neben der beschriebenen Äquivalenz zu einer kausalen Domäne – weitere Vorteile, die sie geeignet machen, kausales Lernen und Denken zu untersuchen. Wie bereits von Steyvers et al. (2003) ausgeführt, existieren in dieser Domäne keine *A-priori-Biases*, welche die Versuchspersonen beim Lösen von Aufgaben nutzen können und damit die Untersuchungsergebnisse verfälschen bzw. deren Interpretierbarkeit erschweren. In anderen, natürlicheren Domänen, die normalerweise eingesetzt werden, um menschliches Kausaldenken zu erforschen – z.B. Viren-Symptom-Beziehungen (wie im einführenden Beispiel in Abschnitt 2.4), Krankheiten und Substanzen im Blut von Tieren, Lebensmittel und Allergien etc. – haben die Versuchspersonen i.d.R. viel Vorwissen, Erfahrungen und mehr oder weniger konkrete Annahmen über die Basisraten der involvierten Ereignisse, die zugrunde liegenden kausalen Mechanismen, die Stärke und die Richtung von Kausalbeziehungen (z.B. ist die kausale Rolle von Krankheiten [Ursachen] und Symptomen [Effekte] vordefiniert, während jedes Alien als Effekt wie auch als Ursache dienen kann, was u.a. eine Manipulation der Kausalrichtung unter Konstanthaltung aller anderen Variablen vereinfacht bzw. überhaupt erst ermöglicht). Während dieses Vorwissen der Versuchspersonen und dessen Interaktion mit den instruierten Eigenschaften des experimentellen Szenarios in anderen Studien keine große Rolle spielen mag, so stellt es bei der Nutzung natürlicher Szenarien jedoch eine Herausforderung dar, wenn man die Annahmen über die zugrunde liegenden

---

Kausalprozesse unkonfundiert manipulieren möchte. Dies ist daher ein weiterer Grund auf die artifizielle Domäne „Gedanken lesender Aliens“ auszuweichen.

In Anlehnung an die Aufgabe, wie sie Rehder und Burnett (2005) verwendet haben, um die Intuitionen bezüglich der Markov-Bedingung zu untersuchen, wurden die Probanden in den folgenden Experimenten in ein Szenario mit vier Außerirdischen – Gonz, Murks, Brxxx und Zoohng – eingeführt. Die Aliens denken dabei jeder entweder an nichts oder an „POR“ (Essen auf außerirdisch). Es wurde verdeutlicht, dass die „POR“-Gedanken von Gonz probabilistisch die „POR“-Gedanken der anderen Aliens verursachen, aber alle Aliens auch für sich manchmal an „POR“ denken. Die entscheidende experimentelle Manipulation bezog sich nun auf die zugrunde liegenden kausalen Prozesse: In der „*Sending*“-Bedingung teilten wir den Versuchspersonen mit, dass Gonz die Fähigkeit habe, seine „POR“-Gedanken auszusenden und in die Köpfe von Murks, Brxxx und Zoohng zu übertragen. In der „*Reading*“-Bedingung wurde den Probanden dagegen gesagt, dass Murks, Brxxx und Zoohng die Gedanken von Gonz lesen können. Die zugrunde liegenden stochastischen Abhängigkeiten für ein Ursache-Effekt-Paar sind in beiden Fällen identisch: Was Murks, Brxxx bzw. Zoohng gerade denken, hängt davon ab, was Gonz gerade denkt, unabhängig, ob Gonz seine Gedanken aussendet oder Murks, Brxxx und Zoohng die Gedanken von Gonz lesen. Gonz ist also statistisch gesehen Ursache und Murks, Brxxx und Zoohng sind Effekte eben dieser Ursache. Interessant ist nun, wie sich die Manipulation des zugrunde liegenden kausalen Mechanismus auf die Einschätzung der Wahrscheinlichkeit eines Zieleffekts, der Aufgabe von Rehder und Burnett (2005) entsprechend, auswirkt. Der interessanteste Testfall ist hierbei offenkundig eine anwesende Ursache (Gonz denkt an „POR“) gepaart mit nur als abwesend beobachteten weiteren Effekten dieser Ursache. In der „*Sending*“-Bedingung wird man die Abwesenheit der Effekte vornehmlich auf einen Fehler des Senders, also der Ursache, attribuieren. Da ein solcher Fehler der Ursache auch weitere Zieleffekte beeinflusst, sollte die Einschätzung der Wahrscheinlichkeit der Anwesenheit des (nicht beobachteten) Zieleffekts im Vergleich zu einer Situation mit nur anwesenden Effekten relativ niedrig ausfallen. In der „*Reading*“-Bedingung hingegen sollte die Abwesenheit beider Effekte eher auf einen Fehler der Gedanken

lesenden Aliens attribuiert werden und daher die Einschätzung der Wahrscheinlichkeit der Anwesenheit des Zieleffekts nicht zu stark von der Anzahl als abwesend beobachteter weiterer Effekte abhängig sein. Im Falle der Abwesenheit der Ursache (Gonz denkt an nichts) sollte der Einfluss der anderen Effekte über beide Bedingungen hinweg jedoch weitestgehend verschwinden, da bei Abwesenheit der Ursache die Effekte nur vom Hintergrund verursacht werden. Annahmen über kausale Mechanismen, wie die gemeinsame Ursache ihre Effekte hervorbringt, sollten daher bei Abwesenheit der Ursache keinen Einfluss auf die Bewertungen des Zieleffekts haben (wenn die Ursache abwesend ist, verursacht sie auch nichts).

### 3.1 Experiment 1

Wie eben beschrieben soll in Experiment 1 empirisch getestet werden, inwieweit Einschätzungen hinsichtlich der Anwesenheit eines Zieleffekts gegeben des Status der anderen Variablen in einer *Common-Cause*-Struktur von den Annahmen der Versuchspersonen über die zugrunde liegenden Kausalprozesse beeinflusst werden; hier mithin durch eine Manipulation des einer Gedankenübertragung bei Außerirdischen zugrunde liegenden Mechanismus, der entweder eine effektseitige Fehlerattribution („*Reading*“) oder eine ursacheseitige Fehlerattribution nahe legt („*Sending*“).

#### 3.1.1 Methode

*Versuchspersonen.* An der Untersuchung nahmen 50 Personen teil (31 Frauen, 19 Männer; Durchschnittsalter 25.6 Jahre, Altersspanne 18–75 Jahre). Diese wurden auf dem Campus, im Zentralen Hörsaalgebäude (ZHG), im Psychologischen Institut, in den Mensen rund um den Campus der Universität Göttingen und zum Teil im Bekannten- und Verwandtenkreis der Versuchsleiterinnen und Versuchsleiter angeworben. Es handelte sich dabei größtenteils um Studierende der Universität Göttingen verschiedenster Fachrichtungen. Die Teilnahme erfolgte freiwillig. Nach Ende der Untersuchung wurden die Teilnehmerinnen und Teilnehmer mit einer Süßigkeit belohnt und sofern gewünscht über den Zweck und das Ziel der Untersuchung genauer aufgeklärt. Die Probanden wurden den beiden Bedingungen (Gedankenlesen vs. Gedankensenden) zufällig zugeteilt ( $N = 2 \times 25$ ).



*Versuchsdurchführung und Material.* Das Experiment bestand aus zwei Phasen, einer Instruktionsphase und einer Testphase. In der *Instruktionsphase* wurde den Versuchspersonen eine siebenseitige Instruktion vorgelegt. In dieser wurde erklärt, dass es vier Aliens gebe, Gonz, Murks, Brxxx und Zoohng (siehe Abbildung 7), die entweder an nichts denken oder an „POR“ (Essen auf außerirdisch). Des Weiteren wurde dargelegt, dass Murks, Brxxx und Zoohng häufig dann an „POR“ denken würden, wenn Gonz an „POR“ denke, dass dies aber nicht immer der Fall sei. Überdies würden die Aliens auch jeder für sich manchmal an Essen („POR“) denken, unabhängig davon, was Gonz gerade denke. Wie oben bereits dargelegt, war die entscheidende Manipulation, welcher kausale Prozess den Relationen zugrunde liegt. In der „*Sending*“-Bedingung wurde den Probanden gesagt<sup>22</sup>, Gonz könne seine „POR“-Gedanken aussenden und damit in die Köpfe von Murks, Brxxx und Zoohng übertragen. Allerdings sei Gonz nicht perfekt, sondern manchmal auch unaufmerksam, so dass er es nicht immer schaffe, seine Gedanken auch auszusenden. In der „*Reading*“-Bedingung hingegen wurde den Versuchspersonen erklärt, dass Murks, Brxxx und Zoohng die „POR“-Gedanken von Gonz lesen könnten. Allerdings seien Murks, Brxxx und Zoohng nicht perfekt, sondern manchmal auch unaufmerksam, und ein unaufmerksames Aliens schaffe es nicht, den Gedanken von Gonz zu lesen.



*Abbildung 7.* Die Abbildungen der vier Außerirdischen Gonz, Murks, Brxxx und Zoohng (verkleinerte Darstellung), wie sie in der Instruktion von Experiment 1 eingeführt wurden. Das Material wurde größtenteils von Steyvers et al. (2003) übernommen.

Zur Verdeutlichung des Gesagten wurden am Beispiel der Außerirdischen Gonz und Murks die Bedeutung der vier möglichen Fälle des an nichts und an

<sup>22</sup> Die im Folgenden und auch in den Beschreibungen der weiteren Experimente durch die Verwendung des Konjunktiv I dargestellte indirekte Rede stellt die Instruktion in ihrem Wortlaut und keine verkürzende Zusammenfassung dar.

„POR“ Denkens genauer erklärt (siehe Abbildung 8): 1) *Beide Aliens denken an nichts*. Da die Außerirdischen nicht immer an „POR“ denken würden, könne es natürlich auch einmal sein, dass beide an nichts denken. 2) *Gonz und Murks denken beide an „POR“*. Murks könne zwar auch aus freien Beweggründen an „POR“ denken, viel wahrscheinlicher sei aber, dass Murks den Gedanken von Gonz gelesen habe („*Reading*“-Bedingung) bzw. dass Gonz seinen Gedanken ausgesendet und damit in den Kopf von Murks übertragen habe („*Sending*“-Bedingung) und Murks darum an „POR“ denke. 3) *Nur Gonz denkt an „POR“*. Gonz denke gerade von selbst an „POR“. Murks sei gerade un aufmerksam und verpasse es sozusagen, den Gedanken von Gonz zu lesen („*Reading*“-Bedingung) bzw. Gonz sei aber nicht konzentriert genug und schaffe es daher nicht, seinen Gedanken auszusenden und damit in den Kopf von Murks zu übertragen („*Sending*“-Bedingung). 4) *Nur Murks denkt an „POR“*. Gonz denke an nichts. Da Murks in diesem Fall auch keinen Gedanken lesen könne („*Reading*“-Bedingung) bzw. Gonz in diesem Fall auch keinen Gedanken aussende und in den Kopf von Murks übertragen könne („*Sending*“-Bedingung), denke Murks aus freien Beweggründen an nichts.

**Fall 1:** Beide denken an nichts.



**Fall 2:** Beide denken an „POR“.



**Fall 3:** Nur Gonz denkt an „POR“.



**Fall 4:** Nur Murks denkt an „POR“.



Abbildung 8. Die vier möglichen Fälle, anhand derer die Instruktion erklärt wurde (deutlich verkleinerte Darstellung). Erläuterungen siehe Text.

---

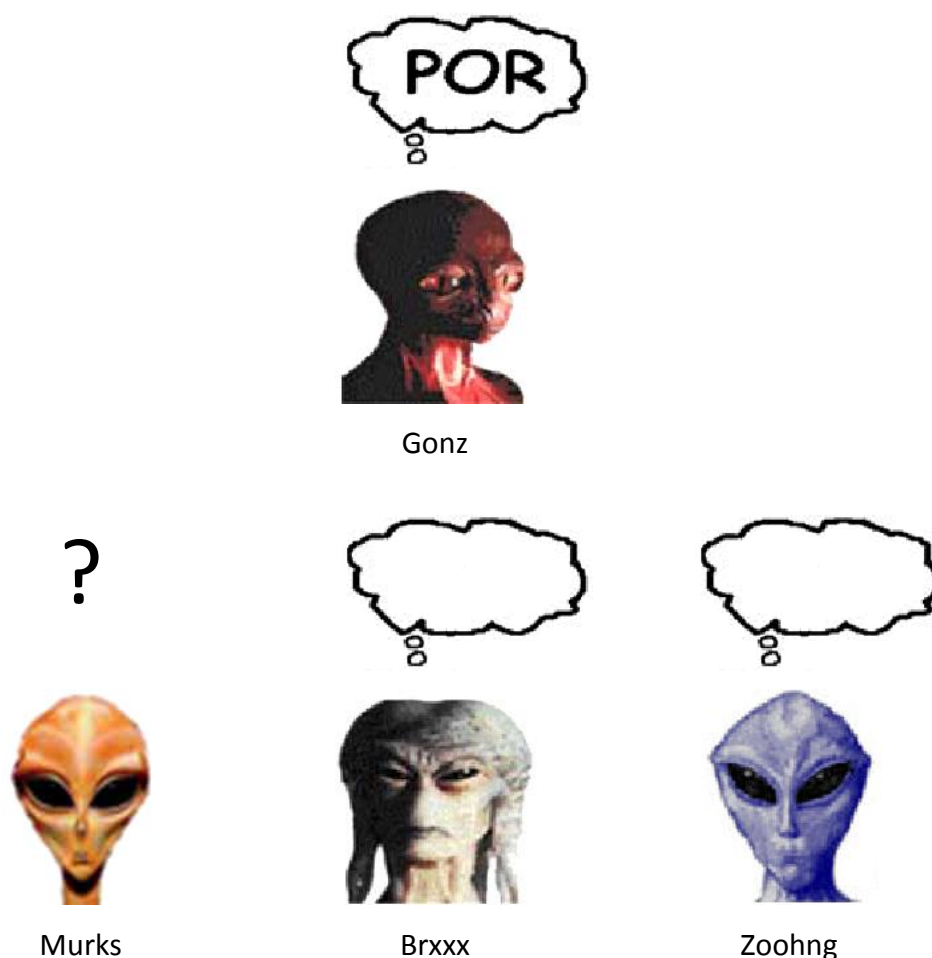
Es wurde verdeutlicht, dass die dargestellten Beziehungen in gleicher Weise auch für Gonz und Brxxx sowie für Gonz und Zoohng gelten würden. Des Weiteren wurde den Versuchspersonen – in Anlehnung an Rehder und Burnett (2005) – erklärt, dass die vier Aliens, wenn sie sich in einem Raum aufhielten, jeweils ca. 70% der Zeit an Essen und ca. 30% der Zeit an nichts denken würden. Die Information über diese Randwahrscheinlichkeiten der Variablen wurde dabei vor allem aus Gründen der Vergleichbarkeit zu den Experimenten von Rehder und Burnett (2005) aufgenommen, die diese Informationen enthielten. Sie sollten den Versuchspersonen primär als Ausgangspunkt für ihre Bewertungen dienen, da die sonstigen Beschreibungen von Kausalstärke und Basisrate nur verbal (wie „manchmal“ oder „häufig“) vorlagen. (Man bemerke, eine Lernphase, in der die Probanden die Parameter des zugrundeliegenden kausalen Systems hätten lernen können, gab es in diesem Experiment, wie auch in allen weiteren Experimenten, nicht.)

Vor Abschluss der Instruktionsphase, nachdem die Versuchspersonen die Instruktion gelesen hatten, wurden die Probanden aufgefordert, den Inhalt der Instruktion noch einmal zu erklären. Dies sollte sicherstellen, dass auch alle wesentlichen Instruktionsteile richtig verstanden wurden. Konnte eine Probandin bzw. ein Proband den Inhalt der Instruktion in seinen wesentlichen Teilen nicht wiedergeben, so musste sie bzw. er diese noch einmal lesen.

In der sich anschließenden *Testphase* wurden den Versuchspersonen Testfälle gezeigt, in denen der Gedanke von Gonz und die Gedanken zweier weiterer Aliens präsentiert wurden (für einen Beispiel-Testfall siehe Abbildung 9). Die Aufgabe bestand nun darin, anzugeben, in wie vielen von zehn vorgestellten Situationen mit dieser Konfiguration das Zielalien (gekennzeichnet mit einem Fragezeichen über dem Kopf) an „POR“ denken werde. Die Angabe erfolgte damit auf einer Rating-Skala von 0 bis 10. Die Einschätzung der Wahrscheinlichkeit der Anwesenheit des Zieleffekts auf einer Häufigkeitsskala anstelle einer Wahrscheinlichkeitsskala wurde gewählt, da Versuchspersonen mit Häufigkeitsskalen in der Regel besser umgehen können. Verzerrende Effekte, die auf eine möglichen Misskonzeption von Wahrscheinlichkeiten zurückgehen, können so vermieden werden (siehe diesbezüglich u.a. die Debatte über den Einsatz von Häufig-

keitsskalen beim *Base-Rate-Neglect*: Brase, 2002; Gigerenzer & Hofrage, 1995, 1999; Lewis & Keren, 1999; Mellers & McGraw, 1999)<sup>23</sup>.

Insgesamt hat jede Versuchsperson sechs Testfälle in zufälliger Reihenfolge gesehen. Sie waren dergestalt konstruiert, dass je Status der gemeinsamen Ursache (hier also die Gedanken von Gonz: nichts vs. „POR“) drei Fälle abgedeckt wurden: Keines der anderen Effektaliens denkt an „POR“, eins von den anderen Effektaliens denkt an „POR“ und beide anderen Effektaliens denken an „POR“. Welche Effektaliens als Ziel dienten, wurde zufällig bestimmt.



*Abbildung 9.* Ein Beispiel-Testfall, wie er in der Testphase von Experiment 1 benutzt wurde. Die Versuchspersonen hatten sich zehn Situationen mit der gegebenen Konfiguration vorzustellen und dann anzugeben, in wie vielen dieser Situationen das Zielalien (hier Murks) an „POR“ denken würde (Skala von 0 bis 10).

<sup>23</sup> Dennoch wird hier terminologisch das relative Häufigkeitsmaß als Wahrscheinlichkeitsmaß interpretiert und die Begriffe „Häufigkeitseinschätzung“ und „Wahrscheinlichkeitsschätzung“ als austauschbar und im Folgenden im Sinne der beschriebenen relativen Häufigkeit verwendet.

Für die Durchführung des Versuchs wurde ein Aktenordner angelegt, der die Instruktionen (je sieben Seiten für die „*Reading*“- bzw. „*Sending*“-Bedingung) und die 24 möglichen Testfälle<sup>24</sup> sowie die Daten- und eine mit Excel erzeugte Randomisierungstabelle enthielt. In der Testphase hat der Versuchsleiter den Versuchspersonen der Randomisierungstabelle entsprechend jeweils die sechs Testfälle nacheinander vorgelegt und die Antwort der Versuchsperson in der Datentabelle dokumentiert. Die Durchführung selbst erfolgte auf dem Campus der Universität Göttingen sowie umliegenden Universitätsgebäuden und dauerte ca. 15 Minuten pro Versuchsperson.

*Design.* Aus dem beschriebenen Procedere ergibt sich ein dreifaktorielles varianzanalytisches 2x2x3-Design mit zwei Messwiederholungsfaktoren. Die interindividuelle Manipulation „*Reading*“ vs. „*Sending*“ (Faktor *RS*, zweistufig) wird ergänzt durch die intraindividuelle Variation des Zustandes der Ursache (Gedanke von Gonz: nichts vs. „*POR*“; Faktor *C*, zweistufig) und der Variation der Anzahl der als anwesend beobachteten Effekte (0, 1 oder 2; Faktor  $N_{E+}$ , dreistufig).

*Vorhersage.* Bereits aus der Bayes-Netz-Theorie ergibt sich die Vorhersage eines deutlichen Haupteffekts bezüglich des Faktors *C*, also dem Zustand der Ursache. Im Falle der Anwesenheit der Ursache (Gonz denkt an „*POR*“) sollte die eingeschätzte Wahrscheinlichkeit der Anwesenheit des Zieleffekts (Zielalien denkt „*POR*“) deutlich höher ausfallen, als bei Abwesenheit der Ursache (Gonz denkt an nichts). Aus den in der Einleitung gemachten Überlegungen im Hinblick auf die „*Reading*“-vs.-„*Sending*“-Manipulation (Faktor *RS*) ergibt sich des Weiteren die Vorhersage einer gerichteten bedingten Interaktion des *RS*-Faktors mit der Anzahl der als anwesend beobachteten weiteren Effekte (Faktor  $N_{E+}$ ) im Falle der Anwesenheit der Ursache dergestalt, dass die Einschätzung umso stärker abfällt je mehr weitere Effekte als abwesend beobachtet werden und dies im „*Sending*“-Fall deutlich ausgeprägter als im „*Reading*“-Fall. Keine bedingte Interaktion jedoch wird zwischen diesen beiden Faktoren im Falle der Abwesenheit der Ursa-

---

<sup>24</sup> Je Zustand der Ursache (nichts vs. „*POR*“) drei Testfälle für „kein Effekalien denkt an Essen“, sechs Testfälle für „ein Effekalien denkt an Essen“ (die Position des an Essen denkenden Aliens kann zusätzlich je Zielalien binär variiert werden) und drei Testfälle für „beide Effekaliens denken an Essen“; also 12 Testfälle je Zustand der Ursache, insgesamt demnach 24.

che erwartet. Aus letzteren beiden Vorhersagen ergibt sich überdies die Vorhersage einer Drei-Wege-Interaktion zwischen den drei Faktoren.

### 3.1.2 Ergebnisse

Wie aus Tabelle 1 sowie Abbildung 10 gut ersichtlich, bestätigen sich die Erwartungen (detaillierte Ergebnisse der varianzanalytischen Auswertung finden sich in Tabelle A-1, Tabelle A-2 und Tabelle A-3 in Anhang A).

Die Häufigkeitseinschätzungen hinsichtlich der Anwesenheit des Zieleffekts sind im Durchschnitt viel höher, wenn die gemeinsame Ursache anwesend, als wenn diese abwesend ist (Haupteffekt C:  $F_{1,48}=111.85$ ,  $p<.001$ ,  $\eta^2=.70$ )<sup>25</sup>.

Im Falle der Anwesenheit der Ursache (Gonz denkt an „POR“; die oberen beiden Linien in Abbildung 10) hängt die Wahrscheinlichkeitseinschätzung bezüglich der Anwesenheit des Zieleffekts deutlich von der Anzahl der weiteren als anwesend („POR“) beobachteten Effekte ab (Haupteffekt  $N_{E+}$  für C=1:  $F_{2,96}=35.80$ ,  $p<.001$ ,  $\eta^2=.43$ ). Ausgehend von null als anwesend beobachteter weiterer Effekte steigt sie mit deren zunehmender Anzahl in der „Reading“-Bedingung von 7.00 auf 8.36 ( $F_{2,48}=11.89$ ,  $p<.001$ ,  $\eta^2=.23$ ) und in der „Sending“-Bedingung von 4.28 auf 8.48 ( $F_{2,48}=29.14$ ,  $p<.001$ ,  $\eta^2=.55$ ). Der jeweils offensichtlich monotone Anstieg ist damit in der „Sending“-Bedingung deutlich ausgeprägter als in der „Reading“-Bedingung (Interaktionseffekt  $RS \times N_{E+}$  für C=1:  $F_{2,96}=10.07$ ,  $p<.001$ ,  $\eta^2=.17$ ).

Im Falle der Abwesenheit der Ursache (Gonz denkt an nichts; die unteren beiden Linien in Abbildung 10) schwanken die Einschätzungen der Wahrscheinlichkeit der Anwesenheit des Zieleffekts relativ unsystematisch zwischen 2.56 und 3.88 und liegen für die „Reading“-Bedingung etwas unter der „Sending“-Bedingung. Weder die Anzahl der als anwesend beobachteten Effekte hat einen signifikanten Einfluss auf die Beurteilungen (Haupteffekt  $N_{E+}$  bei C=0:  $F_{2,96}=1.52$ ,  $p=.22$ )<sup>26</sup> noch deren Interaktion mit dem „Reading“-vs.-„Sending“-Faktor (Interaktionseffekt  $RS \times N_{E+}$  für C=0:  $F_{2,96}=1.32$ ,  $p=.27$ ).

Die Drei-Wege-Interaktion verfehlt knapp die Signifikanzgrenze (Interaktionseffekt  $C \times RS \times N_{E+}$ :  $F_{2,96}=2.97$ ,  $p=.06$ ,  $\eta^2=.06$ ).

<sup>25</sup> Mit  $\eta^2$  sei hier und im Folgenden die partielle varianzanalytische Effektgröße gemeint.

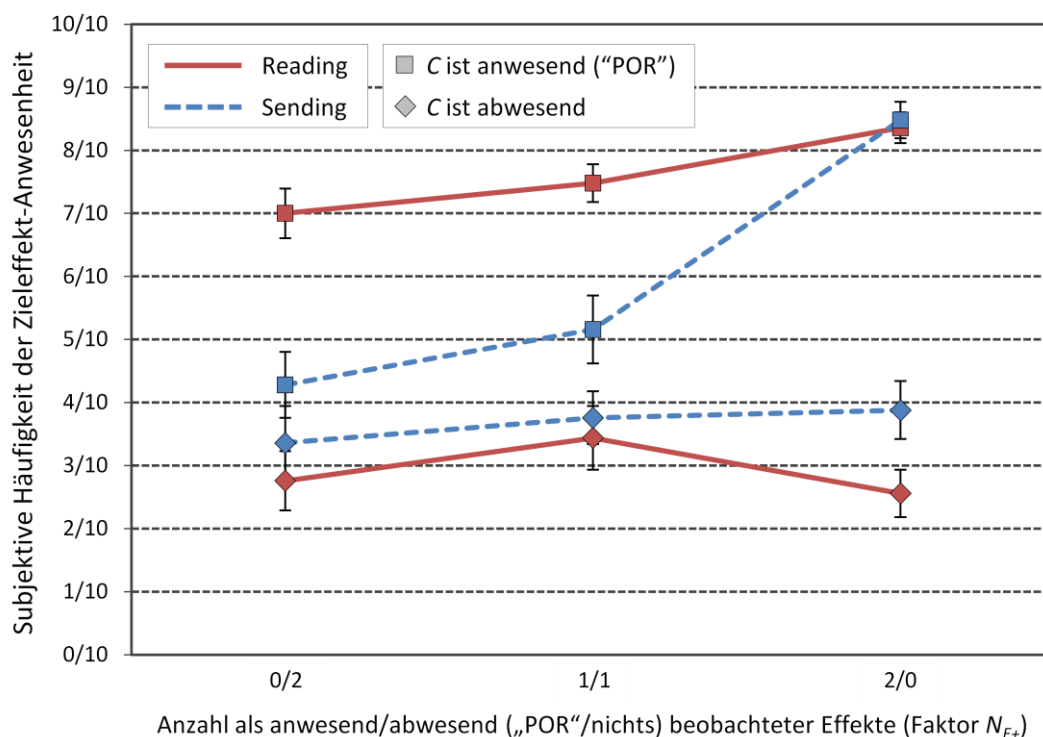
<sup>26</sup> Das gilt sowohl für die „reading“-Bedingung ( $F_{2,48}=1.62$ ,  $p=.21$ ) als auch für die „sending“-Bedingung ( $F_{2,48}=1.05$ ,  $p=.36$ ).

Tabelle 1

## Durchschnittliche Bewertung der Anwesenheit des Zieleffekts (Experiment 1)

	„Reading“-Bedingung (n = 25)			„Sending“-Bedingung (n = 25)		
	$N_{E+} = 0$	$N_{E+} = 1$	$N_{E+} = 2$	$N_{E+} = 0$	$N_{E+} = 1$	$N_{E+} = 2$
C = 1 („POR“)	7.00 (1.96)	7.48 (1.50)	8.36 (1.22)	4.28 (2.62)	5.16 (2.69)	8.48 (1.45)
C = 0 (nichts)	2.76 (2.35)	3.44 (2.53)	2.56 (1.87)	3.36 (2.93)	3.76 (2.09)	3.88 (2.30)

*Anmerkungen.* Die Tabelle zeigt die durchschnittlichen Bewertungen auf einer Skala von „0 von 10“ bis „10 von 10“ (Häufigkeitsskala mit 11 Stufen) bezüglich der Wahrscheinlichkeit der Anwesenheit des Zieleffekts (Standardabweichung in Klammern). C = Zustand der gemeinsamen Ursache;  $N_{E+}$  = Anzahl der als anwesend beobachteten weiteren Effekte.



*Abbildung 10.* Durchschnittliche Bewertungen der Anwesenheit des Zieleffekts in zehn fiktiven Situationen gegeben des Zustandes der gemeinsamen Ursache (Faktor C) und der Anzahl der als anwesend/abwesend beobachteten weiteren Effekte (Faktor  $N_{E+}$ ) in der „Reading“- und „Sending“-Bedingung ( $N = 2 \times 25$ ) von Experiment 1. Die Fehlerbalken entsprechen dem Standardfehler des Mittelwerts.

### 3.1.3 Diskussion

Experiment 1 demonstriert eindrucksvoll, wie stark sich Annahmen der Versuchspersonen hinsichtlich der zugrunde liegenden Kausalprozesse auf die Einschätzung der Anwesenheit eines nicht beobachteten Effekts auswirken können. In der „*Sending*“-Bedingung, deren Beschreibung eine Attribution möglicher Fehler auf die Ursache nahelegt, sind die Abhängigkeiten deutlich stärker als in der „*Reading*“-Bedingung, deren Beschreibung eher eine Attribution auf die Effekte befördert.

Die Erklärungsversuche von Rehder und Burnett (2005) sind durch die Befunde widerlegt. Sowohl das *Feature-Uncertainty*-Modell, wie auch das *Underlying-Mechanism*-Modell sind mit den Ergebnissen nicht kompatibel. Beide Modelle induzieren die Markov-Verletzung durch einen Mechanismus, der unabhängig von der kausalen Rolle der Variablen und auch unabhängig vom betrachteten kausalen System ist: Das *Feature-Uncertainty*-Modell durch die Annahme, dass Unsicherheit um die wahre Ausprägung der involvierten Variablen besteht und das *Underlying-Mechanism*-Modell durch die Annahme eines generativen Mechanismus der Kategorie, der Exemplare mit einheitlichen Ausprägungen produziert, und der dadurch mit der Kausalinferenz konkurriert (siehe Näheres dazu auch Abschnitt 2.3). Beide Modelle sagen symmetrische Antwortmuster hinsichtlich des Zustandes der Ursache vorher (und nicht wie gefunden, unterschiedliche Abhängigkeiten für die Abwesenheit wie die Anwesenheit der Ursache). Überdies sind beide Modelle nicht geeignet, unterschiedliche Annahmen über die zugrunde liegenden Kausalprozesse abzubilden, da sie gleichsam nicht auf diese Bezug nehmen. Damit dürfte eine Manipulation der Annahmen über die zugrunde liegenden kausalen Prozesse keinen Einfluss auf die Wahrscheinlichkeitsschätzungen haben. Das Gegenteil jedoch ist der Fall.

Dennoch bleibt offen, inwieweit eventuell das Ergebnis durch die Instruktion der Randwahrscheinlichkeiten beeinflusst ist. Wie in der Versuchsbeschreibung dargestellt, wurden alle relevanten Parameter des Systems verbal beschrieben („häufig“, „manchmal“ etc.), nur die Randwahrscheinlichkeit, mit der ein Alien an „POR“ denkt, wurde numerisch mit 70% angegeben. Dies erfolgte in Anlehnung an die Studien von Rehder und Burnett (2005), die diese Information



ebenfalls gaben. Es könnte jedoch sein, dass dies zusammen mit der intraindividuellen Manipulation der Faktoren  $C$  und  $N_{E+}$  zu einer Aufforderungscharakteristik dahingehend geführt hat, dass die Versuchspersonen bei ihren Antworten versucht haben, die Randhäufigkeiten herzustellen und sie dabei Markov-Verletzungen erzeugt haben. Eine Erklärung für den Unterschied zwischen der „Reading“- und der „Sending“-Bedingung ist dies zwar nicht, überprüfenswert erscheint es dennoch.

Des Weiteren stellt sich die Frage, ob die verbale Beschreibung des kausalen Systems ein nicht Markov-konformes Antwortverhalten nahe gelegt hat. In der Beschreibung wurde erklärt, dass Murks, Brxxx und Zoohng häufig dann an „POR“ denken, wenn Gonz an „POR“ denkt. Eine Interpretationsmöglichkeit wäre, dass Murks, Brxxx und Zoohng häufig gemeinsam an „POR“ denken, wenn Gonz an „POR“ denkt, oder seltener aber dennoch gemeinsam nicht an „POR“ denken. Diese Interpretation legt ein kausales System nahe, in dem die Markov-Bedingung nicht gilt, ohne dass die Interpretation etwas mit den eigentlichen Annahmen über die zugrunde liegenden Kausalprozesse zu tun hätte. Zwar kann sich auch diese Erklärung nicht wirklich für einen Unterschied zwischen der „Reading“- und der „Sending“-Bedingung verantwortlich zeichnen, dennoch erscheint es sinnvoll zu überprüfen, ob der grundsätzliche Befund davon unabhängig ist. Dies soll im Folgenden Hauptziel von Experiment 2 sein.

## 3.2 Experiment 2

Wie eben dargestellt sollen in Experiment 2 die Befunde von Experiment 1 mit einer leicht geänderten Instruktion repliziert werden, um zu überprüfen, inwieweit diese von der Vorgabe der Randwahrscheinlichkeiten und von der Art der Beschreibung der Parameter des Systems abhängig sind. Dazu soll zum einen auf die Vorgabe der Randwahrscheinlichkeiten verzichtet werden und zum anderen die Parameter nicht mehr auf der Ebene des Gesamtsystems, sondern auf Einzel-link-Ebene instruiert werden.

### 3.2.1 Methode

*Versuchspersonen.* 50 Personen nahmen an der Untersuchung teil (31 Frauen, 19 Männer; Durchschnittsalter 24.6 Jahre, Alterspanne 16–51 Jahre). Die Anwer-

bung erfolgte im Zentralen Hörsaalgebäude (ZHG), im Psychologischen Institut, auf dem Campus, den daran angrenzenden Mensen der Universität Göttingen und zum Teil im Bekannten- und Verwandtenkreis der Versuchsleiterinnen und Versuchsleiter. Es handelte sich dabei größtenteils um Studierende der Universität Göttingen verschiedenster Fachrichtungen. Keiner der Probanden hatte bereits an Experiment 1 teilgenommen. Die Versuchspersonen nahmen freiwillig an dem Experiment teil. Nach Abschluss des Experiments wurden die Teilnehmerinnen und Teilnehmer sofern gewünscht über den Zweck und das Ziel der Untersuchung genauer aufgeklärt und mit einer kleinen Süßigkeit belohnt.

Die Zuteilung der Probanden zu den beiden experimentellen Bedingungen (Gedankenlesen vs. Gedankensenden) erfolgte zufällig ( $N = 2 \times 25$ ).

*Versuchsdurchführung und Material.* Die Durchführung von Experiment 2 entsprach im Wesentlichen der von Experiment 1, d.h. das Experiment bestand aus einer Instruktionsphase und einer Testphase; das Bildmaterial wurde aus Experiment 1 übernommen. In der *Instruktionsphase* wurde den Probanden eine sechsseitige Instruktion vorgelegt, die eine gekürzte Fassung der Instruktion von Experiment 1 darstellt. Insbesondere wurden die Informationen über die Randhäufigkeiten der Variablen entfernt. Überdies wurden die Parameter nicht mehr auf der Ebene des Gesamtsystems beschrieben – wie in Experiment 1 –, sondern am Beispiel der gegebenen Zweierrelation Gonz–Murks, anhand derer auch die vier Fälle demonstriert wurden. Es hieß daher, dass Murks häufig dann an „POR“ denke, wenn Gonz an „POR“ denke, und dass dies für Gonz und Brxxx sowie für Gonz und Zoohng ebenso gelte. Was dies jedoch in der Gesamtbetrachtung aller vier Aliens hieß, wurde damit offen gelassen. Die weiteren Inhalte der Instruktion sind identisch mit denen in Abschnitt 3.1.1 für Experiment 1 beschriebenen. Ebenso wie in Experiment 1 wurden die Versuchspersonen am Ende der Instruktionsphase nach dem Lesen der Instruktion aufgefordert, die wesentlichen Inhalte zu erläutern.

Die *Testphase* war mit Experiment 1 identisch. Den Versuchspersonen wurden insgesamt sechs Testfälle in zufälliger Reihenfolge gezeigt. Die Abbildungen der Testfälle wurden von Experiment 1 übernommen. Es wurde lediglich eine neue Randomisierungstabelle erstellt.

Für die Durchführung wurde ebenso ein Aktenordner erstellt, der die Instruktionen, die Testfälle sowie die Daten- und Randomisierungstabelle enthielt. Die Durchführung erfolgte auf dem Campus der Universität Göttingen und umliegenden Gebäuden. Sie dauerte ca. 15 Minuten pro Versuchsperson.

*Design und Vorhersage.* Design und Vorhersage entsprechen Experiment 1.

### 3.2.2 Ergebnisse

Die Ergebnisse von Experiment 2 sind in Tabelle 2 und Abbildung 11 dargestellt, detaillierte Ergebnisse der varianzanalytischen Auswertung finden sich überdies in Tabelle B-1, Tabelle B-2 und Tabelle B-3 in Anhang B. Ein Blick auf die Abbildung zeigt, dass die Ergebnisse von Experiment 1 im Wesentlichen gut repliziert werden konnten.

Wie in Experiment 1 sind die Wahrscheinlichkeitsbewertungen bezüglich der Anwesenheit des Zieleffekts viel höher, wenn die gemeinsame Ursache anwesend, als wenn diese abwesend ist (Haupteffekt C:  $F_{1,48}=85.91$ ,  $p<.001$ ,  $\eta^2=.64$ ).

Tabelle 2

*Durchschnittliche Bewertung der Anwesenheit des Zieleffekts (Experiment 2)*

	„Reading“-Bedingung (n = 25)			„Sending“-Bedingung (n = 25)		
	$N_{E+} = 0$	$N_{E+} = 1$	$N_{E+} = 2$	$N_{E+} = 0$	$N_{E+} = 1$	$N_{E+} = 2$
C = 1 („POR“)	5.64 (2.04)	6.16 (2.12)	6.68 (2.87)	4.44 (2.63)	5.36 (2.20)	8.04 (2.15)
C = 0 (nichts)	3.16 (1.99)	3.76 (1.90)	4.00 (2.38)	3.28 (2.28)	3.32 (1.60)	4.44 (2.26)

*Anmerkungen.* Die Tabelle zeigt die durchschnittlichen Bewertungen auf einer Skala von „0 von 10“ bis „10 von 10“ (Häufigkeitsskala mit 11 Stufen) bezüglich der Wahrscheinlichkeit der Anwesenheit des Zieleffekts (Standardabweichung in Klammern). C = Zustand der gemeinsamen Ursache;  $N_{E+}$  = Anzahl der als anwesend beobachteten weiteren Effekte.

Im Falle der Anwesenheit der Ursache (Gonz denkt an „POR“; die beiden oberen Linien in Abbildung 11) wird die Einschätzung der Häufigkeit der Anwesenheit des Zieleffekts stark von der Anzahl weiterer als anwesend beobachteter Effekte beeinflusst (Haupteffekt  $N_{E+}$  für C=1:  $F_{2,96}=22.10$ ,  $p<.001$ ,  $\eta^2=.32$ ). Ausge-

hend von null als anwesend beobachteter weiterer Effekte steigt sie mit deren zunehmender Anzahl in der „Reading“-Bedingung leicht von 5.64 auf 6.68 an ( $F_{2,48}=2.62$ ,  $p=.08$ ,  $\eta^2=.10$ ) und liegt damit vom Niveau her deutlich niedriger als in Experiment 1. In der „Sending“-Bedingung steigen sie erheblich von 4.44 auf 8.04 ( $F_{2,48}=23.01$ ,  $p<.001$ ,  $\eta^2=.49$ ) und liegen auf mit Experiment 1 vergleichbarem Niveau. Der Anstieg ist damit in der „Sending“-Bedingung viel stärker ausgeprägt als in der „Reading“-Bedingung (Interaktionseffekt  $RS \times N_{E+}$  für  $C=1$ :  $F_{2,96}=7.43$ ,  $p<.01$ ,  $\eta^2=.13$ ).

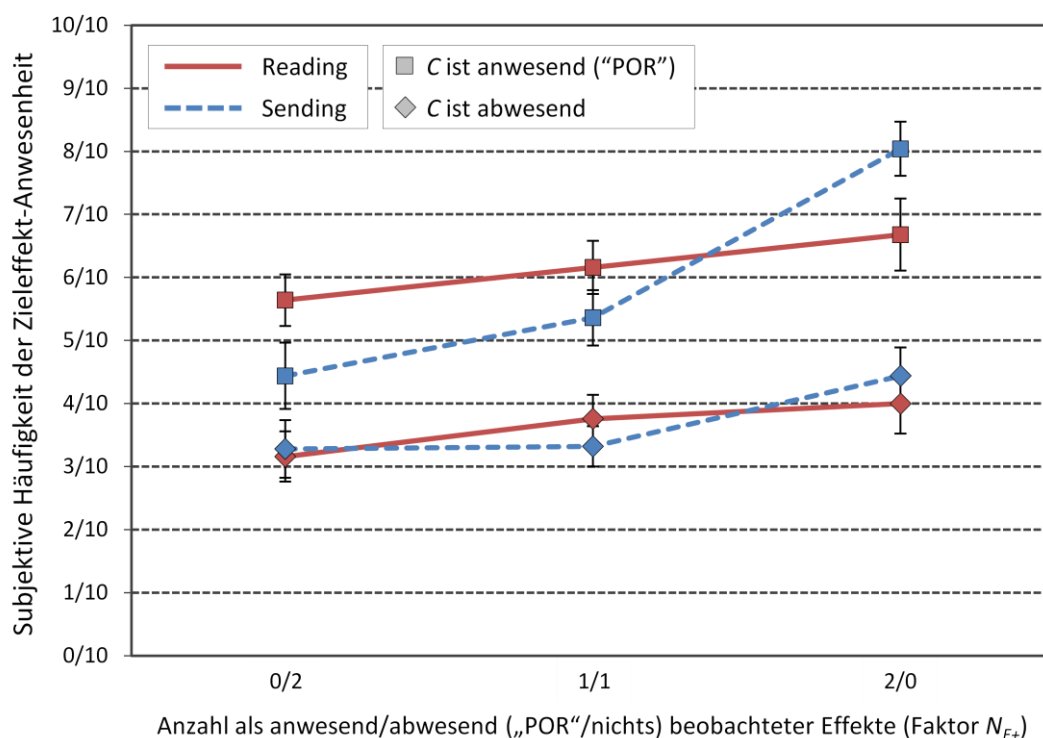


Abbildung 11. Durchschnittliche Bewertungen der Anwesenheit des Zieleffekts in zehn fiktiven Situationen gegeben des Zustandes der gemeinsamen Ursache (Faktor C) und der Anzahl der als anwesend/abwesend beobachteten weiteren Effekte (Faktor  $N_{E+}$ ) in der „Reading“- und „Sending“-Bedingung ( $N = 2 \times 25$ ) von Experiment 2. Die Fehlerbalken entsprechen dem Standardfehler des Mittelwerts.

Im Falle der Abwesenheit der Ursache (Gonz denkt an nichts; die unteren beiden Linien in Abbildung 11) steigen die Bewertungen über beide Bedingungen hinweg gesehen von knapp über 3 bis knapp über 4 signifikant leicht an (Haupteffekt  $N_{E+}$  für  $C=0$ :  $F_{2,96}=4.34$ ,  $p<.05$ ,  $\eta^2=.08$ ), wobei dies vornehmlich auf den deutlicheren Anstieg in der „Sending“-Bedingung zurückzuführen ist ( $F_{2,48}=3.39$ ,  $p<.05$ ,  $\eta^2=.12$ ) und weniger auf die „Reading“-Bedingung, die aber auch einen

leichten, aber nicht signifikanten Anstieg verzeichnet ( $F_{2,48}=1.67$ ,  $p=.20$ ,  $\eta^2=.07$ ). Eine signifikante Interaktion zwischen dem „Reading“-vs.-„Sending“-Faktor und der Anzahl der weiteren als anwesend beobachteten Effekte (Faktor  $N_{E+}$ ) gibt es dennoch nicht (Interaktionseffekt  $RS \times N_{E+}$  für  $C=0$ :  $F_{2,96}<1$ ,  $p=.44$ ), wie vorhergesagt.

Die vorhergesagte Drei-Wege-Interaktion ist signifikant (Interaktionseffekt  $C \times RS \times N_{E+}$ :  $F_{2,96}=3.17$ ,  $p<.05$ ,  $\eta^2=.06$ ); das heißt, dass der Faktor  $RS$  mithin auf beiden Stufen des Faktors  $C$  unterschiedlich auf die Abhängigkeit der Bewertungen von der Anzahl der als anwesend beobachteten weiteren Effekte wirkt.

### 3.2.3 Diskussion

In Experiment 2 konnten die Ergebnisse von Experiment 1 noch einmal gut repliziert werden. In der „Reading“-Bedingung waren die Einschätzungen bezüglich der Anwesenheit des Zieleffekts deutlich weniger vom Status der anderen Effekte abhängig als in der „Sending“-Bedingung. Die konditionale Abhängigkeit der Zielinferenz wird damit also stark von den Annahmen über die zugrundeliegenden Kausalprozesse beeinflusst und zwar unterschiedlich im Falle der Anwesenheit wie im Falle der Abwesenheit der Ursache. Damit erweist sich die Vorhersage auch unter der abgeschwächten Instruktion als stabil, d.h. bei Verzicht der Angabe von Randwahrscheinlichkeiten und der Beschreibung der kausalen Parameter auf Einzellinkebene.

Problematisch erscheint allerdings, dass nunmehr auch im Falle der Abwesenheit der Ursache ein signifikanter Unterschied über die Anzahl der als anwesend beobachteten weiteren Effekte beobachtet werden konnte. Dieser Einfluss der weiteren Effekte ist unabhängig von der experimentellen Manipulation, da eine Interaktion nicht nachgewiesen werden konnte. Berichte von einigen wenigen Probanden deuten darauf hin, dass der Grund für diese Abhängigkeit Annahmen über mögliche weitere externe Ursachen sein könnten, die sich im Falle der Abwesenheit der Ursache für ein gemeinsames Auftreten der Effekte verantwortlich zeichnet (so z.B. die Vermutung, dass die Effektlis alle gerade deshalb an „POR“ – also Essen – denken, weil gerade Mittagszeit ist oder weil sie sich alle in einem Restaurant befinden). Ein solch weiterer generativer externer Einflussfaktor würde sich dabei unabhängig und additiv zusätzlich zu den Variab-

---

len im betrachteten kausalen System auswirken. Die Vorhersagen, die sich ja im Wesentlichen auf Interaktionen mit dem „*Reading*“-vs.-„*Sending*“-Faktor beziehen, werden davon nicht berührt. Zumal scheint der Einfluss von seiner Größe her im Vergleich zu den anderen gefundenen Effekten vernachlässigbar (in der Gesamtdiskussion in Kapitel 7 wird dennoch auf ihn zurückzukommen sein). Das wesentliche Interesse im Folgenden gilt der Untersuchung des Einflusses von Annahmen über die zugrundeliegenden Kausalprozesse.

Im Hinblick auf die Replikation der Ergebnisse in Experiment 2 kann man diesbezüglich einwenden, dass in Experiment 1 so wie auch in Experiment 2 nicht primär die Intuitionen über Kausalprozesse untersucht wurden. Vielmehr wurden die Annahmen relativ direkt durch die in der Instruktion zur Verfügung gestellten Erklärungen manipuliert, indem explizit Gründe für das Versagen – mithin die fehlende Aufmerksamkeit des Senders bzw. der Leser – dargelegt wurden. Die Kausalprozesse – Gedankenlesen und Gedankensenden – wurden daher konkret auch anhand möglicher beteiligter Fehlerprozesse expliziert. Ungeklärt ist daher noch, inwieweit Versuchspersonen solche Annahmen über zugrundeliegende Fehlerstrukturen *a priori* mitbringen und diese Annahmen durch angenommene Kausalprozesse verursacht sind. Dieser Frage soll in Experiment 3 nachgegangen werden.

### **3.3 Experiment 3**

Wie eben dargelegt, soll Experiment 3 zeigen, ob die bisherigen Befunde über den Einfluss von Annahmen über zugrundeliegende Kausalprozesse nicht lediglich dadurch zustande kamen, dass die damit verbundenen Annahmen über Fehlerursachen explizit instruiert wurden. Es könnte daher sein, dass Versuchspersonen nicht automatisch sensitiv für die impliziten Konsequenzen zugrunde liegender Mechanismen sind, sondern nur, wenn ihnen entsprechende Attributionen explizit nahegelegt wurden. Daher soll im Folgenden die experimentelle Manipulation über die zugrunde liegenden Kausalprozesse in ihrer vagest möglichen Form vermittelt werden – ohne Bezug auf eventuelle Ursachen für Fehler. Hauptziel des Experiments ist daher zu klären, ob die Probanden die in Experiment 1 und Experiment 2 explizit instruierte Fehlerstruktur auch selbst aus den

beschriebenen Kausalprozessen ableiten und der Effekt (im Falle seines Auftretens) auf entsprechende Fehlerattributionen zurückgeführt werden kann.

### 3.3.1 Methode

*Versuchspersonen.* 60 Studierende der Universität Göttingen nahmen an der Untersuchung teil (44 Frauen, 16 Männer; Durchschnittsalter 23.9 Jahre, Altersspanne 20–34 Jahre). Die Anwerbung erfolgte im Zentralen Hörsaalgebäude (ZHG) der Universität, im Psychologischen Institut, auf dem Campus, den daran angrenzenden Mensen der Universität Göttingen und zum Teil im Bekanntenkreis der Versuchsleiterinnen und Versuchsleiter. Die Teilnehmerinnen und Teilnehmer rekrutierten sich aus den verschiedensten Fachrichtungen. Keiner der Probanden hatte bereits an einem der vorhergehenden Experimente der Serie teilgenommen. Die Versuchspersonen nahmen freiwillig an dem Experiment teil. Nach Abschluss des Experiments wurden die Teilnehmerinnen und Teilnehmer sofern gewünscht über den Zweck und das Ziel der Untersuchung genauer aufgeklärt und mit einer kleinen Süßigkeit belohnt.

Die Probanden wurden zufällig auf die beiden experimentellen Bedingungen (Gedankenlesen vs. Gedankensenden) aufgeteilt ( $N = 2 \times 30$ ).

*Versuchsdurchführung und Material.* Wie in den beiden vorhergehenden Experimenten bestand Experiment 3 aus zwei Phasen, der Instruktionsphase und der Testphase. In der *Instruktionsphase* wurde den Probanden eine stark gekürzte, dreiseitige Fassung der Instruktion vorgelegt, die keine Hinweise mehr zu eventuellen Fehlerursachen enthielt. Wieder wurde den Versuchspersonen mitgeteilt, dass es vier Aliens – Gonz, Murks, Brxxx und Zoohng – gebe, die meistens an nichts und manchmal an „POR“ denken würden (Bildmaterial siehe Abbildung 12). Des Weiteren wurde erklärt, Gonz könne seine „POR“-Gedanken in die Köpfe von Murks, Brxxx und Zoohng übertragen („*Sending*“-Bedingung) bzw. dass Murks, Brxxx und Zoohng die „POR“-Gedanken von Gonz lesen könnten („*Reading*“-Bedingung). Im Konkreten heiße dies, dass Murks, Brxxx und Zoohng häufig dann an „POR“ denken, wenn Gonz an „POR“ denke, und dass Murks, Brxxx und Zoohng auch jeder für sich manchmal an „POR“ denken, unabhängig davon, was Gonz gerade denke.

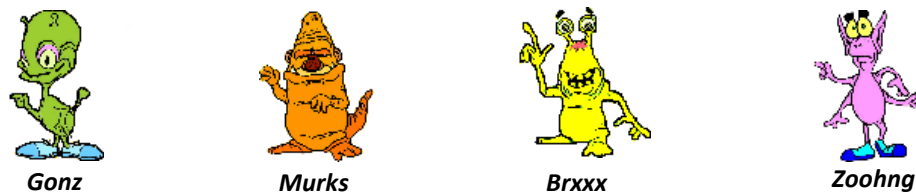


Abbildung 12. Die Abbildungen der vier Außerirdischen Gonz, Murks, Brxxx und Zoohng, wie sie in der Instruktion von Experiment 3 eingeführt wurden. Das Material wurde größtenteils von Steyvers et al. (2003) übernommen, grafisch aber überarbeitet<sup>27</sup>.

Am Beispiel von Gonz und Murks wurden wieder die vier möglichen Fälle zur Verdeutlichung des Gesagten erklärt, unter der Prämisse, dass diese auch für Gonz und Brxxx sowie Gonz und Zoohng gelten würden, und (im Unterschied zu den beiden vorangegangenen Experimenten) unter Verzicht auf Hinweise hinsichtlich möglicher Fehlerquellen: 1) *Gonz und Murks denken beide an nichts*. Da die Aliens nur manchmal an „POR“ denken, sei es recht häufig, dass beide gleichzeitig an nichts denken. Gonz und Murks würden also jeweils zufällig an nichts denken. 2) *Gonz denkt an „POR“ und Murks denkt auch an „POR“*. Da Murks die Gedanken von Gonz meistens lese („*Reading*“-Bedingung) bzw. da Gonz seine Gedanken meistens in den Kopf von Murks übertrage („*Sending*“-Bedingung), sei das recht häufig der Fall. 3) *Gonz denkt an „POR“ und Murks denkt an nichts*. Da Murks meistens die Gedanken von Gonz lese („*Reading*“-Bedingung) bzw. Gonz meistens seine Gedanken in den Kopf von Murks übertrage („*Sending*“-Bedingung), sei dies recht selten der Fall. 4) *Gonz denkt an nichts und Murks denkt an „POR“*. Da nichts nicht gelesen („*Reading*“-Bedingung) bzw. übertragen („*Sending*“-Bedingung) werden könne, denke Murks aus freien Beweggründen an „POR“. Dies sei aber recht selten der Fall.

Vor Abschluss der Instruktionsphase, nach dem Lesen der Versuchsinstruktion, sollten die Probanden die Instruktion noch einmal in ihren wesentlichen Inhalten wiedergeben, um diese dann ggf. nochmals zu lesen.

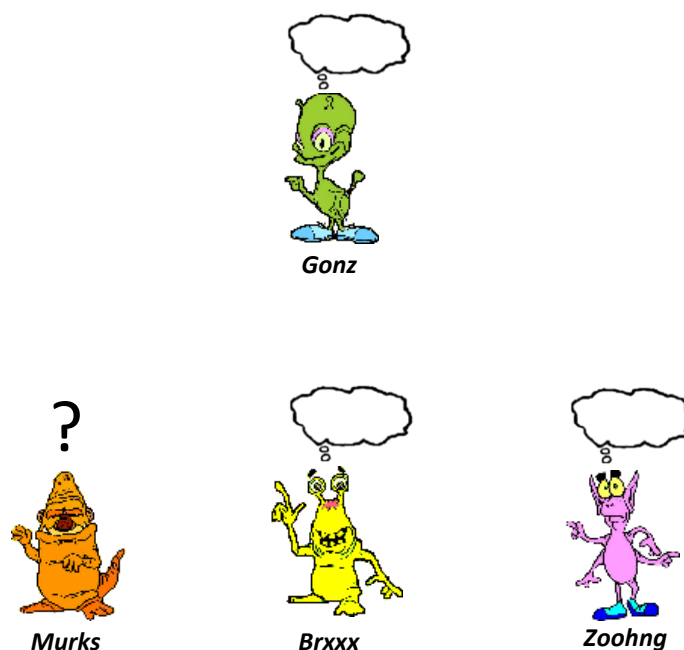
In der *Testphase* wurden den Versuchspersonen wieder je sechs Testfälle in randomisierter Reihenfolge vorgelegt, in denen sie für eine vorgegebene Konfi-

<sup>27</sup> Eine Überarbeitung erfolgte im Hinblick auf die Farben der Aliens dahingehend, dass sich diese alle möglichst gleich unähnlich sehen.



guration die Anwesenheit des Zieleffekts auf einer Skala von 0 bis 10 einschätzen sollten (für weitere Details siehe Experiment 1; ein Beispieltestfall mit dem neuen Bildmaterial findet sich in Abbildung 13). Es handelte sich um die gleichen Testfälle wie in Experiment 1 und 2.

Nachdem die Versuchspersonen die Standardtestfälle bearbeitet hatten, wurde ihnen eine Zusatzfrage vorgelegt. Dieser enthielt nochmals die Konfiguration, in der Gonz an „POR“ und zwei Effektlis, z.B. Brxxx und Zoohng, an nichts dachten (das verbleibende Effektlis wurde in Korrespondenz mit den ursprünglichen Testfällen mit einem Fragezeichen markiert). Ihnen wurde mitgeteilt, dass hier irgendetwas nicht zu stimmen scheine, und dann gefragt, ob dies ihrer Meinung nach eher an Gonz oder eher an Brxxx und Zoohng liege. Die Bewertung wurde auf einer fünfstufigen Ratingskala von -2 (eher an Gonz) bis +2 (eher an Brxxx/Zoohng) erfasst. Welche der Effektlis an nichts dachten, wurde über die Versuchspersonen randomisiert.



*Abbildung 13.* Ein Beispiel-Testfall, wie er in der Testphase von Experiment 3 benutzt wurde. Die Versuchspersonen hatten sich zehn Situationen mit der gegebenen Konfiguration vorzustellen und dann anzugeben, in wie vielen dieser Situationen das Zielalien (hier Murks) an „POR“ denken würde (Skala von 0 bis 10).

Für die Durchführung des Experiments wurde wie in Experiment 1 und Experiment 2 ein Aktenordner erstellt, der die Instruktionen, die Testfälle sowie die

Daten- und eine Randomisierungstabelle enthielt. Die Durchführung erfolgte auf dem Campus der Universität Göttingen und in umliegenden Gebäuden. Sie dauerte ca. 10–15 Minuten pro Versuchsperson.

*Design.* Es handelt sich wieder um ein 2x2x3-Design mit einem experimentellen Faktor („Reading“ vs. „Sending“, Faktor *RS*) und zwei Messwiederholungsfaktoren (Status der Ursache, Faktor *C*, sowie Anzahl der als anwesend beobachteten weiteren Effekte, Faktor  $N_{E+}$ ). Da die Zusatzfrage nur eine Konfiguration enthielt, wird sie über den experimentellen Faktor als gerichteter Vergleich ausgewertet.

*Vorhersage.* Sofern die Hypothese, dass die Befunde in den beiden vorhergehenden Experimenten wirklich auf eine Manipulation der Annahmen über zugrunde liegende Kausalprozesse zurückzuführen ist und nicht auf eine mehr oder weniger explizit instruierte Fehlerstruktur, sollten sich dieselben Muster zeigen wie in Experiment 1 und Experiment 2. Das heißt, insbesondere im Falle der Anwesenheit der Ursache sollte sich ein deutlicher Einfluss der Anzahl als anwesend beobachteter Effekte in der „Sending“-Bedingung und ein weniger starker Einfluss in der „Reading“-Bedingung zeigen. Überdies sollte es im Falle der Abwesenheit der Ursache keine Interaktion zwischen dem „Reading“-vs.-„Sending“-Faktor und der Anzahl als anwesend beobachteter Effekte geben.

Für die Zusatzfrage, die die Fehlerattribution im Falle des Versagens der Ursache abbildet, sollte sich in der „Sending“-Bedingung eine Attribution eher auf *Gonz*, d.h. niedrigere Werte auf der Skala von -2 bis +2, und in der „Reading“-Bedingung eher auf *Brxxx/Zoohng*, d.h. höhere Werte, ergeben. Die Differenz zwischen der „Reading“- und der „Sending“-Bedingung sollte mithin positiv sein.

### 3.3.2 Ergebnisse

Die Ergebnisse für den ersten Teil von Experiment 3 finden sich in Tabelle 3 und Abbildung 14; die detaillierte varianzanalytische Auswertung überdies in Tabelle C-1, Tabelle C-2 und Tabelle C-3 in Anhang C. Insgesamt konnte der Befund aus den ersten beiden Experimenten nochmals gut repliziert werden.

Die Einschätzungen hinsichtlich der Anwesenheit des Zieleffekts sind im Durchschnitt viel höher, wenn die gemeinsame Ursache anwesend, als wenn sie abwesend ist (Haupteffekt *C*:  $F_{1,58}=208.92$ ,  $p<.001$ ,  $\eta^2=.78$ ).

Ist die Ursache anwesend (Gonz denkt an „POR“; die beiden oberen Linien in Abbildung 14) wird die Einschätzung der Häufigkeit der Anwesenheit des Ziel-effekts sehr stark von der Anzahl weiterer als anwesend beobachteter Effekte beeinflusst (Haupteffekt  $N_{E+}$  für  $C=1$ :  $F_{2,116}=12.29$ ,  $p<.001$ ,  $\eta^2=.17$ ). Sie steigt ausgehend von null als anwesend beobachteter weiterer Effekte mit deren zunehmender Anzahl in der „Reading“-Bedingung nur unwesentlich von 7.10 auf 7.50 an ( $F_{2,58}=1.26$ ,  $p=.29$ ,  $\eta^2=.04$ ). In der „Sending“-Bedingung hingegen steigen die Einschätzungen deutlich von 5.10 auf 7.83 ( $F_{2,58}=13.06$ ,  $p<.001$ ,  $\eta^2=.31$ ), insgesamt allerdings etwas schwächer als in Experiment 1 und in Experiment 2. Der Anstieg ist dennoch in der „Sending“-Bedingung viel stärker ausgeprägt als in der „Reading“-Bedingung (Interaktionseffekt  $RS \times N_{E+}$  für  $C=1$ :  $F_{2,116}=6.93$ ,  $p<.01$ ,  $\eta^2=.10$ ).

Ist die Ursache abwesend (Gonz denkt an nichts; die unteren beiden Linien in Abbildung 14) schwanken die mittleren Einschätzungen relativ unsystematisch, aber dennoch statistisch bedeutend zwischen 2.43 und 3.23 (Haupteffekt  $N_{E+}$  für  $C=0$ :  $F_{2,116}=4.93$ ,  $p<.01$ ,  $\eta^2=.08$ ), wobei dies vornehmlich auf die „Sending“-Bedingung ( $F_{2,58}=4.21$ ,  $p<.05$ ,  $\eta^2=.13$ ) und weniger auf die „Reading“-Bedingung ( $F_{2,58}=1.50$ ,  $p=.23$ ,  $\eta^2=.05$ ) zurückzuführen ist. Eine signifikante Interaktion zwischen dem „Reading“-vs.-„Sending“-Faktor und der Anzahl der weiteren als anwesend beobachteten Effekte (Faktor  $N_{E+}$ ) kann wie vorhergesagt jedoch nicht nachgewiesen werden (Interaktionseffekt  $RS \times N_{E+}$  für  $C=0$ :  $F_{2,96}<1$ ,  $p=.50$ ).

Die vorhergesagte Drei-Wege-Interaktion ist wie erwartet signifikant (Interaktionseffekt  $C \times RS \times N_{E+}$ :  $F_{2,96}=3.19$ ,  $p<.05$ ,  $\eta^2=.05$ ); mithin wirkt der Faktor  $RS$  auf beiden Stufen des Faktors  $C$  unterschiedlich auf die Abhängigkeit der Bewertungen von der Anzahl der als anwesend beobachteten weiteren Effekte.

Bei der Zusatzfrage, also inwieweit das Problem bei einer als anwesend beobachteten Ursache und zwei als abwesend beobachteten Effekte mehr auf die Seite der Ursache (-2) oder mehr auf die Seite der Effekte (+2) zurückzuführen sei (fünfstufige Ratingskala), gaben die Versuchspersonen im Mittel in der „Sending“-Bedingung -0.43 ( $sd = 1.28$ ) und in der „Reading“-Bedingung 1.30 ( $sd = 0.70$ ) an. Damit wurde der Fehler in der „Sending“-Bedingung eher auf die Ursa-

che attribuiert, während er in der „Reading“-Bedingung eher effektseitig gesehen wurde ( $t_{58}=6.51, p<.001. d_{emp}=1.68$ ), was der Vorhersage entspricht.

Tabelle 3

Durchschnittliche Bewertung der Anwesenheit des Zieleffekts (Experiment 3)

	„Reading“-Bedingung (n = 30)			„Sending“-Bedingung (n = 30)		
	$N_{E+} = 0$	$N_{E+} = 1$	$N_{E+} = 2$	$N_{E+} = 0$	$N_{E+} = 1$	$N_{E+} = 2$
C = 1 („POR“)	7.10 (2.48)	6.93 (1.87)	7.50 (1.96)	5.10 (2.55)	6.13 (2.16)	7.83 (1.60)
C = 0 (nichts)	2.80 (1.88)	2.57 (1.50)	3.10 (2.25)	2.43 (1.74)	2.53 (1.36)	3.23 (2.03)

Anmerkungen. Die Tabelle zeigt die durchschnittlichen Bewertungen auf einer Skala von „0 von 10“ bis „10 von 10“ (Häufigkeitsskala mit 11 Stufen) bezüglich der Wahrscheinlichkeit der Anwesenheit des Zieleffekts (Standardabweichung in Klammern). C = Zustand der gemeinsamen Ursache;  $N_{E+}$  = Anzahl der als anwesend beobachteten weiteren Effekte.

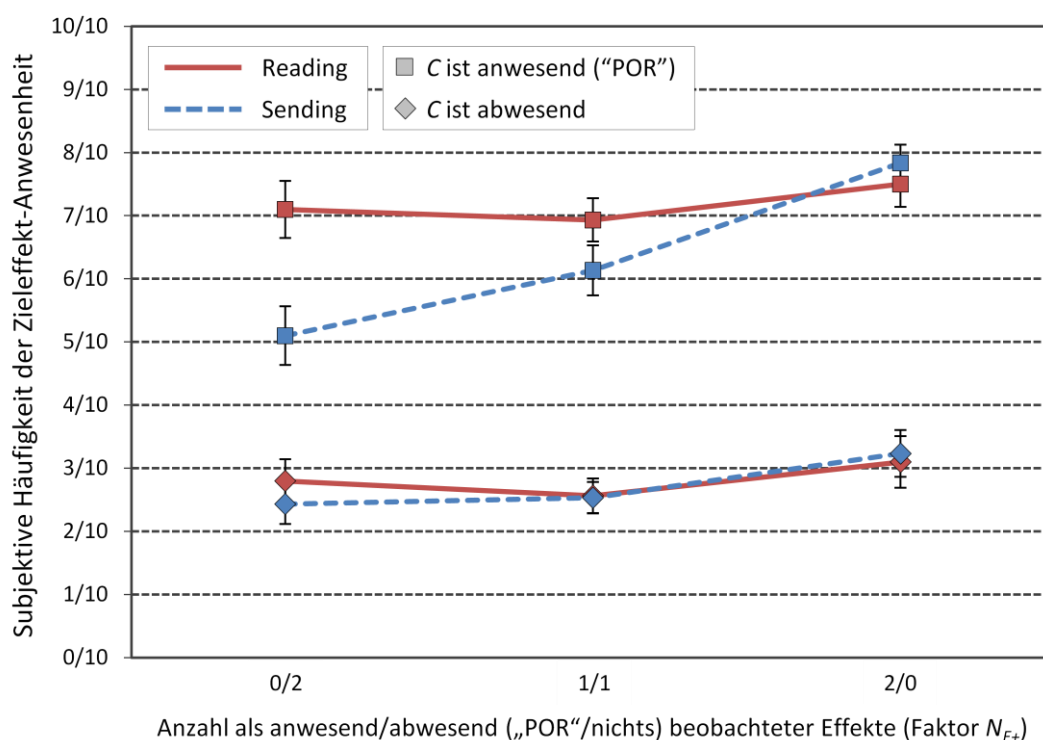


Abbildung 14. Durchschnittliche Bewertungen der Anwesenheit des Zieleffekts in zehn fiktiven Situationen gegeben des Zustandes der gemeinsamen Ursache (Faktor C) und der Anzahl der als anwesend/abwesend beobachteten weiteren Effekte (Faktor  $N_{E+}$ ) in der „Reading“- und „Sending“-Bedingung ( $N = 2 \times 30$ ) von Experiment 3. Die Fehlerbalken entsprechen dem Standardfehler des Mittelwerts.

### 3.3.3 Diskussion

In Experiment 3 konnten mit einer deutlich impliziteren Manipulation der Annahmen der Versuchspersonen über die zugrunde liegenden Kausalprozesse die Ergebnisse von Experiment 1 und von Experiment 2 bestätigt werden. Die Manipulation bezog sich nunmehr nur noch auf kausale Mechanismen (Gedankenlesen vs. Gedankensenden) und nicht auch auf damit zusammenhängende Annahmen über eine mögliche Fehlerstruktur. Dennoch zeigte sich im Falle der Anwesenheit der gemeinsamen Ursache wieder ein starker Einfluss der weiteren beobachteten Effekte (d.h. eine Markov-Verletzung) in der „*Sending*“-Bedingung, nicht jedoch in der „*Reading*“-Bedingung. Dieser Befund scheint dabei auf die gleichen Schlussprozesse zurückzuführen zu sein, wie in den beiden vorhergehenden Experimenten. Zwar mangelte es in der Instruktion an einem Hinweis über mögliche, bedingungsabhängige Fehlerursachen, diese wurden von den Probanden aber in gleicher Weise abgeleitet: In der Zusatzfrage, die am Ende des Experiments eingeführt wurde, um die Fehlerattribution abzufragen, zeigten die Versuchspersonen im Falle der Anwesenheit der Ursache aber der Abwesenheit zweier ihrer Effekte eine starke Tendenz einer Attribution dieser Abwesenheit auf die Ursache in der „*Sending*“-Bedingung und eine Attribution eher auf die Effekte in der „*Reading*“-Bedingung. Damit konnte die grundsätzliche, der Manipulation zugrundeliegende Idee bestätigt werden.

Wie bereits in Experiment 2 wurde auch in Experiment 3 ein Einfluss der Anzahl der als anwesend beobachteten Effekte im Falle der Abwesenheit der Ursache gefunden. Eine Interaktion mit dem „*Reading*“-vs.-„*Sending*“-Faktor konnte allerdings nicht nachgewiesen werden. Damit scheint auch hier die Annahme (durch zumindest einige Probanden) einer zusätzlichen, externen generativen Ursache vorzuliegen, die unabhängig von der experimentellen Manipulation wirkt und diese Abhängigkeiten induziert.

### 3.4 Zusammenfassung

Im vorliegenden Kapitel wurden drei Experimente vorgestellt, die deutlich gezeigt haben, dass Einschätzungen hinsichtlich der Anwesenheit eines Zieleffekts gegeben des Status der anderen Variablen in einer *Common-Cause*-Struktur systematisch von den Annahmen der Versuchspersonen über die zugrunde liegen-

den Kausalprozesse beeinflusst werden. Hierzu wurden die Probanden in ein *Common-Cause*-Szenario auf der Basis der „*Mind-Reader-Alien*“-Aufgabe von Steyvers et al. eingeführt, das es ermöglichte, die Annahmen der Versuchspersonen über die zugrundeliegenden Prozesse experimentell zu manipulieren. Je nachdem, ob die gemeinsame Ursache als Gedankensender („*Sending*“-Bedingung) oder die Effekte als Gedankenleser („*Reading*“-Bedingung) beschrieben wurden, waren die Versuchspersonen bei der Beurteilung der Anwesenheit eines nicht beobachteten Effekts gegeben der Anwesenheit der gemeinsamen Ursache unterschiedlich sensitiv für die Anzahl weiterer als anwesend beobachteter Effekte dieser gemeinsamen Ursache. In der „*Sending*“-Bedingung stiegen die Beurteilungen hinsichtlich der Anwesenheit des Zieffekts sehr stark mit steigender Anzahl weiterer als anwesend beobachteter Effekte an, während die Abhängigkeit in der „*Reading*“-Bedingung deutlich geringer ausgeprägt war. Im Fall der Abwesenheit der Ursache waren die Bewertungen zwar ebenfalls teilweise sensitiv für die Anzahl der als anwesend beobachteten Effekte, jedoch unabhängig von der experimentellen Manipulation hinsichtlich der zugrunde liegenden Kausalprozesse.

Dieses insgesamt eigentlich der Markov-Bedingung widersprechende Bewertungsverhalten deutet darauf hin, dass Probanden bei der Lösung dieser Art von Aufgaben deutlich mehr Wissen – insbesondere im Hinblick auf die Konsequenzen der den Kausalrelationen zugrunde liegenden Mechanismen – berücksichtigen als bisher angenommen. In Experiment 3 wurde explizit getestet, ob diese zusätzlichen Annahmen bezüglich der Kausalprozesse sich in unterschiedlichen Annahmen über die zugrunde liegende Fehlerstruktur des kausalen Systems äußern. Im Fall der Anwesenheit der Ursache, jedoch der Abwesenheit aller ihrer beobachteten Effekte, attribuierten die Versuchspersonen dieses Problem in der „*Sending*“-Bedingung auf die Ursache, in der „*Reading*“-Bedingung jedoch auf die Effekte.

## 4 Modellierung

Wie im letzten Kapitel dargelegt, nutzen Versuchspersonen bei der Beurteilung der Anwesenheit eines unbeobachteten Effekts in einer *Common-Cause*-Struktur gegeben des Status der gemeinsamen Ursache und der weiteren Effekte mehr Vorwissen, als eine einfache Analyse auf Bayes-Netz-Ebene suggeriert. Die Ratings im Hinblick auf die Anwesenheit des unbeobachteten Effekts hängen stark von den weiteren Eigenschaften des instruierten Systems ab, vor allem von den Annahmen über die zugrunde liegenden Kausalprozesse und die implizierte Fehlerstruktur.

Dies mag auch nicht verwundern. Wie bereits in Kapitel 2 beschrieben, setzen Bayes-Netze gerade aufgrund der Markov-Bedingung voraus, dass die kausalen Prozesse, die von einer Ursache ausgehen, völlig unabhängig voneinander sind. Das heißt, ist die Ursache anwesend, dann bestimmt sich für jeden ihrer Effekte jeweils unabhängig, ob sie „ursächlich“ für einen Effekt wird. Dass die Ursache jedoch in Gänze versagen könnte und daher keinen ihrer Effekte hervorbringt, ist nicht vorgesehen<sup>28</sup>. Auf statistisch-computationaler Ebene stellt dies zwar kein Problem dar: Die Markov-Bedingung ist eine Bedingung, die hergestellt werden kann, indem weitere Variablen in das betrachtete System eingeführt werden, die die konditionalen Abhängigkeiten erklären (siehe dazu Pearl, 2000, oder auch Hausman & Woodward, 1999), auf psychologischer Ebene jedoch erscheint es unplausibel, als Ausgangspunkt ein kausales System zu definieren, in dem die Markov-Bedingung gilt (also ein klassisches Bayes-Netz zu postulieren). Rehder und Burnett (2005) wie auch die in Kapitel 3 präsentierten Studien zeigen, dass die Markov-Bedingung gerade keine Default-Annahme im menschlichen Kausaldenken darzustellen scheint. Es liegt daher nahe, eine Beschreibungsform zu wählen, die dies explizit berücksichtigt.

Im vorliegenden Abschnitt soll daher ein bayesianisches Modell entwickelt werden, das zum einen – aufbauend auf der in Abschnitt 2.4 entwickelten Idee – klassische Bayes-Netze um eine weitere, gemeinsame Fehlerkomponente erweitert, die gleichermaßen auf alle Effekte einer Ursache wirkt, und zum anderen

---

<sup>28</sup> Wenn man von dem Fall absieht, dass zufällig alle Ursache-Effekt-Relationen unabhängig voneinander versagen.

den stattfindenden Schlussprozess als Bayes-Inferenz modelliert. Aus dem Modell werden sodann konkrete Vorhersagen abgeleitet, die mit den Befunden aus Kapitel 3 verglichen werden können. Des Weiteren werden die qualitativ-strukturellen Modellvorhersagen in einer Analyse in Anlehnung an die Überlegungen von Roberts und Pashler (2000) auf ihre Stabilität hin überprüft, insbesondere inwieweit diese von den freien Parametern abhängig sind.

#### 4.1 Modellierungsidee

Betrachten wir als Ausgangspunkt die Problemstellung, wie sie in Kapitel 3 empirisch untersucht wurde, und das dort zugrunde gelegte, einfache *Common-Cause*-Modell mit einer Ursache  $C$  und mit drei Effekten  $E_1$ ,  $E_2$  und  $E_3$ . In den verwendeten verbal instruierten Szenarien stehen in der Regel Informationen zur Verfügung, aus denen Annahmen über die Basisraten der involvierten Variablen (Basisrate der Ursache,  $b_C$ , und Basisrate der Effekte,  $b_E$ ) sowie über deren Relationen ( $C \rightarrow E_1$ ,  $C \rightarrow E_2$ ,  $C \rightarrow E_3$ ) und deren Kausalstärke ( $w_C$ ) in mehr oder weniger präziser Form<sup>29</sup> abgeleitet werden können. Unter der Markov-Bedingung beschreibt das Komplement der Kausalstärke  $w_C$  die Misserfolgswahrscheinlichkeit für jeden Link unter der Annahme von deren Unabhängigkeit. Sie erlaubt damit, die drei Relationen zu einem Bayes-Netz zusammenzufügen. Der Kausalstärkeparameter  $w_C$  beschreibt also die Stärke, mit der eine Ursache einen ihrer Effekte unabhängig von den anderen hervorbringt, für die es jeweils einen analogen Parameter gibt. Es fehlt demnach ein Parameter, mit dem das kognitive System seine Annahmen über die Stärke der Abhängigkeiten dieser Relationen repräsentieren könnte.

Auf computationaler Ebene kann die Markov-Bedingung – wie oben erläutert (siehe auch Abschnitt 2.2) – durch Hinzufügen weiterer, unbeobachteter Variablen hergestellt werden, die die konditionalen Abhängigkeiten erklären. Eine Hauptursache für konditionale Abhängigkeiten können dabei – wie bereits in Abschnitt 2.4 ausgeführt – unterschiedliche Fehlerquellen im Hinblick auf die kausalen Links darstellen, die wiederum Annahmen über das Zusammenwirken

<sup>29</sup> Das diesbezüglich verwendeten Begriffen wie „häufig“ oder „selten“ eine immanente Unsicherheit im Hinblick auf den durch sie beschriebenen Parameter anhaftet, wird weiter unten dadurch Rechnung getragen, dass über diesen Parameter Wahrscheinlichkeitsverteilungen definiert werden, die diese Unsicherheit abbilden.



der kausalen Mechanismen widerspiegeln. In einem Bayes-Netz werden mögliche Fehlerquellen als unabhängig gedacht: Jedem kausalen Link ist damit eine hypothetische, unabhängige Fehlerquelle zugeordnet, die mit Wahrscheinlichkeit  $(1 - w_C)$  die Ursache daran hindert, den entsprechenden Effekt hervorzubringen. Diese hypothetischen Fehlervariablen sind unbeobachtet. Ihr Einfluss äußert sich im Parameter  $w_C$ . Je schwächer die Fehlerquellen, desto größer  $w_C$ . Es liegt nahe – insbesondere im Hinblick auf die in Kapitel 3 durchgeführten Studien – das kausale System (hier ein *Common-Cause*-Modell) einfach um eine weitere Fehlerquelle zu ergänzen, die auf alle kausalen Links gleichermaßen präventiv wirkt, indem sie deren Kausalstärke mindert. Auch diese Fehlerquelle sei unbeobachtet. Ihre Existenz erzeugt aber auf der Ebene der beobachtbaren Variablen des kausalen Systems konditionale Abhängigkeiten, mithin also Markov-Verletzungen. Je stärker die gemeinsame Fehlerquelle, desto stärker diese Abhängigkeiten. Mit dieser gemeinsamen Fehlerquelle lassen sich somit unterschiedliche Annahmen hinsichtlich der Abhängigkeiten der Kausalrelationen abbilden: Versagen diese eher gemeinsam, ist mithin also eher die Ursache für Fehler verantwortlich (wie z.B. in der „*Sending*“-Bedingung), dann ist die gemeinsame Fehlerquelle stark; sind die kausalen Links dagegen nahezu unabhängig voneinander, also eher die Effekte für Fehler verantwortlich (wie z.B. in der „*Reading*“-Bedingung), dann ist die gemeinsame Fehlerquelle eher schwach.

Abbildung 15 zeigt dieses um die gemeinsame Fehlerquelle *PN* (*preventive noise*) erweiterte *Common-Cause*-Modell mit der Ursache *C*, den Effekten  $E_1, \dots, E_n$  sowie den Parametern  $w_C, w_{PN}, b_C, b_{PN}$  und  $b_E$ , wobei

- $w_C \in [0,1]$  – die Kausalstärke von *C* im Hinblick auf die Effekte,
- $w_{PN} \in [0,1]$  – die präventive Stärke von *PN* im Hinblick auf die Links  $C-E_i$ ,
- $b_C \in [0,1]$  – die Basisrate der gemeinsamen Ursache *C*,
- $b_{PN} \in [0,1]$  – die Basisrate von *PN* (*preventive noise*) und
- $b_E \in [0,1]$  – die Basisrate der Effekte  $E_1, \dots, E_n$ .

Es wird dabei vereinfachend davon ausgegangen, dass die Kausalstärke  $w_C$  für alle Ursache-Effekt-Relationen identisch ist, wie auch alle Effekte die gleiche

Basisrate  $b_E$  haben. Die grundsätzliche Idee der Einführung einer gemeinsamen Fehlerursache bleibt davon aber unberührt und die grundlegenden Modellvorhersagen werden von dieser vereinfachenden Annahme nicht beeinflusst. Sie liegt aber durch die Eigenschaften des instruierten Systems in den beschriebenen Experimenten nahe.

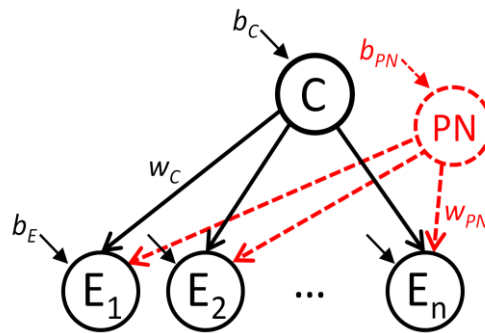


Abbildung 15: Ein um eine gemeinsame Fehlerursache  $PN$  erweitertes *Common-Cause*-Modell mit  $n$  Effekten. Die Kausalstärke  $w_C$  gilt für alle Links  $C-E_i$ , die Basisrate  $b_E$  gilt für alle Effekte und die präventive Stärke  $w_{PN}$  gilt für alle Links  $PN-E_i$ .

Tabelle 4

Wahrscheinlichkeitstabelle für die Anwesenheit eines Effektes  $E_i$  gegeben des Status der gemeinsamen Ursache  $C$  und der gemeinsamen Fehlerquelle  $PN$

$C$	$PN$	$P(E_i = 1 C, PN)$
0	0	$b_E$
0	1	$b_E$
1	0	$b_E + w_C - w_C b_E$
1	1	$b_E + w_C(1 - w_{PN}) - w_C(1 - w_{PN}) b_E$

Anmerkungen.  $b_E$  – Basisrate des Effektes  $E_i$ ,  $w_C$  – Stärke der Ursache  $C$ ,  $w_{PN}$  – Stärke der präventiven Ursache  $PN$ .

Die eingeführte Fehlerursache wirkt nun in folgender Weise: Ist sie abwesend, dann behindert sie die Ursache  $C$  nicht, ihre Effekte mit der Wahrscheinlichkeit  $w_C$  zu produzieren. Ist die gemeinsame Fehlerquelle jedoch anwesend, dann reduziert sie die Wirkung von  $C$  auf die Effekte um den Faktor  $(1 - w_{PN})$ , die Gesamtstärke von  $C$  ist in diesem Fall also  $w_C(1 - w_{PN})$ . Ist  $w_{PN}$  hoch, die gemeinsame Fehlerursache also stark, folgt daraus eine starke Absenkung der Wirkung von  $C$ . Es handelt sich um eine *Noisy-AND-NOT*-Verknüpfung von  $C$  und  $PN$  und

darauf aufbauend um eine *Noisy-OR*-Verknüpfung mit der generativen Hintergrundursache von  $E$  (siehe hierzu Cheng, 1997; Glymour, 1998; Griffiths & Tenenbaum, 2005). Die bedingte Wahrscheinlichkeit eines Effekts lässt sich damit in einer Wahrscheinlichkeitstabelle darstellen (siehe Tabelle 4).

## 4.2 Ableitung des Basismodells

Im Folgenden wird die Zielinferenz, die den Experimenten in Kapitel 3 zugrunde liegt, bayesianisch abgeleitet. Die Darstellung beschränkt sich aus Gründen der Übersichtlichkeit auf die zentralen Ableitungsschritte. Eine ausführliche Ableitung findet sich in Anhang D.

Ausgangspunkt für die Modellierung ist also die Zielfrage, die die Versuchsperson in den Experimenten zu beantworten hatte, mithin die Beurteilung der Wahrscheinlichkeit der Anwesenheit des nicht beobachteten Effekts  $E_n$  gegeben des Status der gemeinsamen Ursache  $C$  und der anderen beobachteten Effekte,  $E_1, \dots, E_{n-1}$ :

$$P(E_n = 1 | C, E_1, \dots, E_{n-1}; \omega) \quad (5)$$

Der Parametervektor  $\omega$  stellt hierbei das abstrakte Vorwissen bzw. die Annahmen der Versuchsperson über die kausale Domäne dar. Mathematisch gesehen handelt es sich um einen Vektor von Hyperparametern, also den Parametern der verschiedenen *A-priori*-Verteilungen der Parameter des repräsentierten kausalen Systems, denn Vorwissen wird in computationalen Bayes-Modellen i.d.R. nicht als Punktschätzer beschrieben (z.B. eine Kausalstärke von .7), sondern als Wahrscheinlichkeitsverteilungen über die interessierenden Parameter (also z.B. eine Betaverteilung über die Kausalstärke mit den Hyperparametern  $\alpha$  und  $\beta$ ), um nicht nur den Wert an sich, sondern auch die an ihm haftende Unsicherheit abzubilden. Die genauen Parameter des kausalen Systems sind dem kognitiven System schließlich immer unbekannt.

Um mit den Parametern dennoch rechnen zu können, führen wir sie in das Modell ein und integrieren sie aus<sup>30</sup>:

<sup>30</sup> Da der Parametervektor  $w$  aus fünf Komponenten besteht, handelt es sich bei dem Integral eigentlich um ein Fünffachintegral. Weil der Parametervektor in der folgenden Ableitung jedoch nicht aufgelöst wird, also immer über den Vektor  $w$  integriert wird, soll das Doppelintegralzeichen andeuten, dass es sich um ein Mehrfachintegral handelt.

$$\begin{aligned}
P(E_n = 1|C, E_1, \dots, E_{n-1}; \omega) \\
= \iint P(E_n = 1|C, E_1, \dots, E_{n-1}; w) \cdot P(w|\omega) dw
\end{aligned} \tag{6}$$

mit

$$w = (w_C, w_{PN}, b_C, b_{PN}, b_E) \tag{7}$$

und unter der Annahme, dass

$$P(w|C, E_1, \dots, E_{n-1}; \omega) = P(w|\omega) \tag{8}$$

Letztere Annahme besagt, dass die Parameter des kausalen Systems unabhängig sind von den beobachteten Zuständen der gemeinsamen Ursache  $C$  und den Effekten  $E_1, \dots, E_{n-1}$  und die Parameter daher vollständig durch die Hyperparameter der *A-priori*-Verteilungen beschrieben sind. Das heißt, es findet kein Lernen statt. Diese Vereinfachung lässt sich rechtfertigen (bzw. sie ist sogar geboten), weil eine experimentelle Situation modelliert wird, in der die präsentierten Testfälle im Hinblick auf ihre Häufigkeit nicht repräsentativ sind für das zugrunde liegende kausale System. Die sechs Testfälle sind – aus experimentellen Gründen – gleichverteilt und daher nicht als i.i.d.-Stichprobe<sup>31</sup> aus der durch das Bayes-Netz beschriebenen gemeinsamen Wahrscheinlichkeitsverteilung anzusehen. Anders zu bewerten wäre eine experimentelle Situation, in der die Probanden fortlaufend unvollständige Stichproben aus dem kausalen System vervollständigen müssten, die dieses auch generiert hat (also ein klassisches Lernexperiment mit Rückmeldung). Dann wären diese Stichproben auch informativ und obige Annahme dürfte nicht getroffen werden.

Hinsichtlich Gleichung 6 stellt sich nun die Frage, wie die bedingte Wahrscheinlichkeit  $P(E_n = 1|C, E_1, \dots, E_{n-1}; w)$  aus der Sicht des oben beschriebenen Modells darzustellen ist. In einem *Common-Cause*-Modell mit Markov-Bedingung würde sich dieser Ausdruck stark vereinfachen:

$$\begin{aligned}
P(E_n = 1|C, E_1, \dots, E_{n-1}; w) &= P(E_n = 1|C; w) \\
&= \begin{cases} b_E & |C = 0 \\ w_C + b_E - w_C b_E & |C = 1 \end{cases}
\end{aligned} \tag{9}$$

---

<sup>31</sup> Eine i.i.d.-Stichprobe ist eine Stichprobe, deren Realisierungen unabhängig voneinander aus der gleichen Wahrscheinlichkeitsverteilung gezogen wurden (i.i.d.: *independent and identically distributed*)

Denn die Inferenz über den Zieleffekt  $E_n$  hängt nicht vom Status der anderen beobachteten Effekte ab, sondern nur vom Status der gemeinsamen Ursache (Markov-Bedingung) und natürlich den Parametern des kausalen Systems, hier mithin von der Basisrate des (Ziel-)Effekts und der Kausalstärke der Ursache.

Der zugehörige Ausdruck für das erweiterte Modell, in dem durch den  $PN$ -Knoten eine gemeinsame Fehlerursache aufgenommen wurde, kann abgeleitet werden durch Aufnahme von  $PN$  als binäre Variable. Da sie unbeobachtet ist, integrieren wir sie durch Summation über die beiden Zustände  $PN = 0$  und  $PN = 1$  aus:

$$\begin{aligned} P(E_n = 1|C, E_1, \dots, E_{n-1}; w) \\ = \sum_{PN} P(E_n = 1|C, PN, E_1, \dots, E_{n-1}; w) \cdot P(PN|C, E_1, \dots, E_{n-1}; w) \end{aligned} \quad (10)$$

Da  $E_n$  gegeben  $C$  und  $PN$  unabhängig von den anderen Effekten ist, die Markov-Bedingung in dem erweiterten Modell also gilt<sup>32</sup>, vereinfacht sich der Ausdruck entsprechend:

$$\begin{aligned} P(E_n = 1|C, E_1, \dots, E_{n-1}; w) \\ = \sum_{PN} P(E_n = 1|C, PN; w) \cdot P(PN|C, E_1, \dots, E_{n-1}; w) \end{aligned} \quad (11)$$

Das Modell lässt sich abstrakt also als Prozess mit zwei Stufen beschreiben: Der erste Schritt beinhaltet die Schätzung, wie wahrscheinlich es ist, dass  $PN$  anwesend bzw. abwesend ist, gegeben der beobachteten Zustände der gemeinsamen Ursache und der Effekte. Ist die Ursache z.B. anwesend, alle ihre Effekte jedoch abwesend, so ist es sehr wahrscheinlich, dass  $PN$  anwesend ist und damit die Wirkung von  $C$  auf ihre Effekte stark vermindert hat. Ist die Ursache jedoch anwesend und so auch ihre Effekte, ist es sehr unwahrscheinlich, dass  $PN$  anwesend ist.

Unter der fiktiven Annahme der Anwesenheit bzw. Abwesenheit der präventiven Ursache  $PN$  lässt sich die Wahrscheinlichkeit der Anwesenheit des Target-Effekts in einem zweiten Schritt leicht bestimmen. Es handelt sich um die bekannte Wahrscheinlichkeitstabelle, die sich aus der Bayes-Netz-Repräsentation

---

<sup>32</sup> Die Markov-Bedingung ist hier also nicht mehr Teil des psychologischen Modells, sondern nur noch auf der computationalen Ebene relevant. Sie dient hier der Vereinfachung der Rechnungen.

des kausalen Systems ergibt. Die eigentliche Zielinferenz  $P(E_n = 1|C, E_1, \dots, E_{n-1}; w)$  bestimmt sich dann als gewichtetes Mittel der Wahrscheinlichkeiten  $P(E_n = 1|C, PN = 0; w)$  und  $P(E_n = 1|C, PN = 1; w)$  – gewichtet mit der Wahrscheinlichkeit von  $PN = 0$  bzw.  $PN = 1$ .

### 4.3 Modellvorhersagen

Bevor aus dem Modell Vorhersagen abgeleitet werden können, müssen Annahmen über die eingehenden Parameter getroffen werden. Diese beschreiben das Wissen der Versuchsperson zum Zeitpunkt der Inferenz. Da es in den modellierten Experimenten jedoch keine Lernphase, sondern lediglich eine Instruktionsphase gibt, in denen die Eigenschaften des kausalen Systems verbal beschrieben werden, müssen die Parameter des Modells auf plausible mit der Instruktion in Einklang stehende Werte gesetzt werden<sup>33</sup>. In den verwendeten Instruktionen wurde beschrieben, dass die Effektaliens häufig dann an „POR“ denken würden, wenn das Ursachealien an „POR“ denkt. Die Kausalstärke ( $w_C$ ) wurde damit als hoch beschrieben. Es liegt daher nahe, diesen Parameter aus folgenden Verteilungen zu ziehen:  $w_C \sim \text{Beta}(10,1)$  mit einem Erwartungswert von  $E[w_C] = 0.91$ . Im Hinblick auf die Effekte wurde des Weiteren instruiert, dass diese manchmal auch von alleine an „POR“ denken würden. Die Basisrate der Effekte ( $b_E$ ) wurde also als relativ niedrig beschrieben; sie soll daher aus folgender Verteilung gezogen werden:  $b_E \sim \text{Beta}(1,2)$  mit einem Erwartungswert von  $E[b_E] = 0.33$ . Über die Basisrate der präventiven Hintergrundursache ( $b_{PN}$ ) wurde keine Aussage getroffen; daher soll diese aus einer nicht-informativen *A-priori*-Verteilung (*flat prior*) gezogen werden:  $b_{PN} \sim \text{Beta}(1,1)$ . Über die Basisrate der Ursache  $b_C$  müssen keine Annahmen getroffen werden, da der Status der Ursache immer beobachtet ist und die Basisrate damit keinen Einfluss auf die Zielinferenz hat.

Der Parameter  $w_{PN}$ , also die präventive Stärke der gemeinsamen Fehlerquelle, beschreibt nun (zusammen mit  $w_C$ ) die Verteilung der Fehler im kausalen

<sup>33</sup> Die Parameter (bzw. die Hyperparameter der entsprechenden Beta-Verteilungen) könnten natürlich auch *post hoc*, also mit Hilfe der Daten aus den Experimenten geschätzt werden. Da das Modell hier vor allem aber im Hinblick auf seine qualitativen Vorhersagen untersucht werden soll, erscheint dieses Vorgehen entbehrlich. In den Abschnitten 4.4 und 4.5 wird auch geprüft, wie die Modellvorhersagen auf Veränderungen der Annahmen reagieren.

System. Ist  $w_{PN}$  hoch, die gemeinsame Fehlerquelle also stark, liegen Fehler eher auf Seite der Ursache und betreffen damit alle Effekte (wie in der „Sending“-Bedingung). Ist  $w_{PN}$  sehr niedrig, dann liegen Fehler eher auf der Seite der Effekte und sie betreffen damit auch nur einzelne Effekte (wie in der „Reading“-Bedingung).

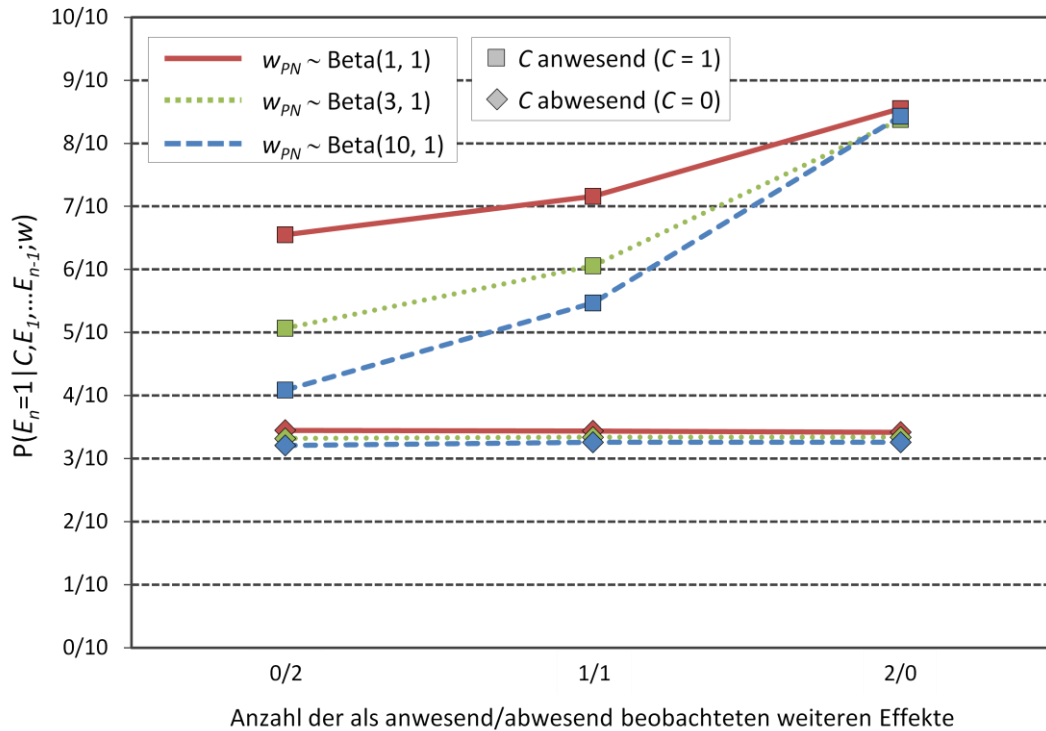


Abbildung 16: Modellvorhersage für unterschiedliche Stärken ( $w_{PN}$ ) der präventiven Ursache  $PN$  auf der Basis einer Monte-Carlo-Simulation mit 100 000 Durchgängen in einem *Common-Cause*-Modell mit einer Ursache  $C$  und drei Effekten. Die weiteren Parameter des kausalen System wurden hierfür aus folgenden Verteilungen gezogen:  $w_C \sim \text{Beta}(10, 1)$ ,  $b_E \sim \text{Beta}(1, 2)$  und  $b_{PN} \sim \text{Beta}(1, 1)$ ;  $b_C$  hat auf die Modellvorhersage keinen Einfluss.

Die Vorhersagen für  $C = 0$  liegen übereinander; die drei Linien wurden daher mit einem Versatz (+0.1, 0, -0.1) versehen, um sie optisch sichtbar zu machen.

Abbildung 16 zeigt die Vorhersage des Modells für die Zielinferenz, also die Wahrscheinlichkeit der Anwesenheit des unbeobachteten Effekts gegeben der Anwesenheit bzw. Abwesenheit der gemeinsamen Ursache und der Anzahl der anwesenden weiteren Effekte der Ursache, in Abhängigkeit von verschiedenen Annahmen über  $w_{PN}$ .  $w_{PN}$  wurde hierfür aus drei verschiedenen Verteilungen gezogen, die unterschiedliche Stärken der gemeinsamen Fehlerursache repräsentieren:  $\text{Beta}(1, 1)$ , also eine nicht informative Verteilung (schwache gemein-

---

same Fehlerquelle),  $Beta(3,1)$  mit einem Erwartungswert von  $E[w_{PN}] = 0.75$  (mittelstarke gemeinsame Fehlerquelle) und  $Beta(10,1)$  mit einem Erwartungswert von  $E[w_{PN}] = 0.91$  (starke gemeinsame Fehlerquelle).

In Abbildung 16 zeigt sich deutlich, wie unterschiedliche Annahmen über die Stärke der gemeinsamen Fehlerursache die Höhe der konditionalen Abhängigen moduliert. Wenn die Ursache  $C$  abwesend ist, dann hat die Stärke der Fehlerursache keinen Einfluss; da  $C$  bei Abwesenheit nicht kausal aktiv ist, ihre Effekte also gerade nicht generiert, kann auch kein Fehler auftreten. Das Modell sagt also im Falle der Abwesenheit von  $C$  (die unteren drei Linien in Abbildung 16) voraus, dass die Einschätzung der Wahrscheinlichkeit der Anwesenheit des Zieleffekts unabhängig sein sollte von a) dem Status der anderen Effekte und b) von den Annahmen über die Stärke der gemeinsamen Fehlerursache. Diese Vorhersage ist in Einklang mit den Ergebnissen der drei in Kapitel 3 vorgestellten Experimente. Die Vorhersage selbst befindet sich auf dem Niveau der Basisrate der Effekte (hier 0.33), da bei Abwesenheit von  $C$  nur der generative Background  $E$  hervorbringt (und die Wahrscheinlichkeit dafür ist gerade die Basisrate  $b_E$ ).

Wenn die gemeinsame Ursache  $C$  anwesend und damit auch kausal aktiv ist (die oberen drei Linien in Abbildung 16), zeigt sich der Einfluss der gemeinsamen Fehlerursache. Je stärker die Fehlerursache, desto stärker der Einfluss des Status der anderen Effekte der gemeinsamen Ursache. Gäbe es die gemeinsame Fehlerursache nicht, gilt also die Markov-Bedingung, dann müsste die Wahrscheinlichkeit der Anwesenheit des Zieleffekts unabhängig sein vom Status der anderen Effekte. Resultat wäre eine horizontale Linie auf hohem Niveau (z.B. ausgehend vom Punkt oben rechts). Nimmt man eine starke Fehlerursache an, dann sollte deren (nicht beobachtbare) Anwesenheit umso wahrscheinlicher werden, je mehr Effekte der Ursache abwesend sind. Ist die gemeinsame Fehlerursache anwesend und stark, dann sollte sie die Ursache  $C$  daran hindern, den Zieleffekt zu generieren. In diesem Fall liegt die Vorhersage damit nur wenig über der Basisrate von  $E$ . Je schwächer die gemeinsame Fehlerursache, desto schwächer ihre präventive Wirkung auf  $C$  und desto geringer sollte der Abfall von rechts nach links ausfallen.



Das Modell sagt damit bei Anwesenheit von  $C$  einen starken Einfluss der Anzahl der anwesenden weiteren Effekte vorher, wenn die Versuchspersonen von einem kausalen System ausgehen, in dem Fehler aufgrund des zugrunde liegenden kausalen Mechanismus eher auf die Ursache zu attribuieren sind (also z.B. Gedankensenden). Liegt jedoch ein System vor, in dem Fehler eher auf die Effekte zu attribuieren sind (z.B. Gedankenlesen), dann sagt das Modell nur einen schwachen Einfluss der Anzahl der weiteren Effekte vorher. Es handelt sich damit um die Vorhersage einer Interaktion zwischen diesen beiden Faktoren und über die Zustände von  $C$  hinweg betrachtet, sogar um eine Drei-Wege-Interaktion (da für die Abwesenheit von  $C$  keine Interaktion zwischen starker/schwacher Fehlerursache und Anzahl anwesender weiterer Effekte vorhergesagt wird). Auch diese Vorhersagen stehen in Einklang mit den Ergebnissen der in Kapitel 3 präsentierten Experimente.

#### 4.4 Sensitivitätsanalysen

Im vorherigen Abschnitt wurde gezeigt, dass das Modell die Daten sehr gut erklären kann und dass der Parameter  $w_{PN}$  die Annahmen über die Abhängigkeit der kausalen Links widerspiegelt. Bevor aus dem Modell (bzw. der grundsätzlichen Idee der Einführung einer korrelierten Fehlerursache) weitere Vorhersagen abgeleitet und getestet werden, soll in diesem Abschnitt gezeigt werden, wie stark die Modellvorhersage von den weiteren Modellparametern abhängig ist. Da das Modell insgesamt fünf Parameter<sup>34</sup> hat ( $w_C$ ,  $w_{PN}$ ,  $b_C$ ,  $b_{PN}$ ,  $b_E$ ), könnte es sein, dass die im letzten Abschnitt und in Abbildung 16 abgeleitete Modellvorhersage stark von den Annahmen über die weiteren Parameter abhängig ist. Nach Roberts und Pashler (2000) hat ein Modell, das so flexibel ist, dass es bei geeigneter Parameterwahl jede beliebige Reihung der Bedingungen vorhersagen kann, keinen Erklärungswert, da es sich an jeden beliebigen Datensatz anpassen lässt. Im Folgenden wird daher gezeigt, inwieweit die qualitativ-strukturelle Vorhersage des Mo-

<sup>34</sup> Die Parameter stellen dabei keine Punktschätzer dar, sondern werden durch Beta-Verteilungen repräsentiert, die ihrerseits wieder jeweils zwei Parameter (Hyperparameter) haben. Da im vorliegenden Fall kein Lernprozess modelliert wird und die Vorhersage lediglich vom Erwartungswert der *A-priori*-Verteilungen abhängt, ist der Erwartungswert der jeweiligen Verteilung der relevante freie Parameter. Die Form der Verteilung spielt nur eine Rolle, wenn Lernen stattfindet. Dies wurde hier jedoch nicht modelliert. Im Folgenden wird daher zwar von Parameter gesprochen; gemeint jedoch ist der Erwartungswert der zugehörigen *A-priori*-Verteilung.

---

dells im Hinblick auf die Wahl der Kausalstärke der Ursache ( $w_C$ ), der Basisrate der Ursache ( $b_C$ ), der Basisrate der Effekte ( $b_E$ ) und der Basisrate der gemeinsamen Fehlerursache ( $w_{PN}$ ) unabhängig ist.

#### 4.4.1 *Abhängigkeit von der Basisrate der Effekte ( $b_E$ )*

Wie aus Tabelle 4 ersichtlich, stellt  $b_E$  eine Untergrenze für die bedingte Wahrscheinlichkeit des Effekts gegeben der gemeinsamen Ursache  $C$  und der gemeinsamen Fehlerquelle  $PN$  dar, weil  $b_E$  in allen vier Kombinationsmöglichkeiten von  $C$  und  $PN$  eine additive Konstante ist. Da die resultierende Modellvorhersage nur ein gewichtetes Mittel ist (siehe Gleichung 11), ist  $b_E$  mithin auch eine Untergrenze der Modellvorhersage. Zusätzlich stellt  $b_E$  (genauer der Erwartungswert von  $b_E$ ) im Falle der Abwesenheit von  $C$  nicht nur die Untergrenze, sondern die Modellvorhersage dar, da  $P(E_n = 1|C = 0, PN) = b_E$  für  $PN = 0$  und  $PN = 1$ .

Dies lässt sich auch in Abbildung 16 erkennen. Zusätzlich wurde in Abbildung 17 die Modellvorhersage für die Ziehung des Parameters  $b_E$  aus einer Beta(1,1)-Verteilung dargestellt, die einen Erwartungswert von 0.5 hat [statt einer Beta(1,2)-Verteilung mit Erwartungswert 0.33 wie in Abbildung 16]. Gut zu sehen ist die Verschiebung der Vorhersage für die Abwesenheit von  $C$  nach oben und die gleichsame Stauchung der Vorhersage für die Anwesenheit von  $C$ . Die grundsätzliche Struktur der Vorhersage – also keine Interaktion bei Abwesenheit von  $C$ , Vorliegen einer Interaktion bei Anwesenheit von  $C$  – bleibt jedoch erhalten, sie ist damit grundsätzlich unabhängig von der Wahl von  $b_E$  (von Extremwerten einmal abgesehen).

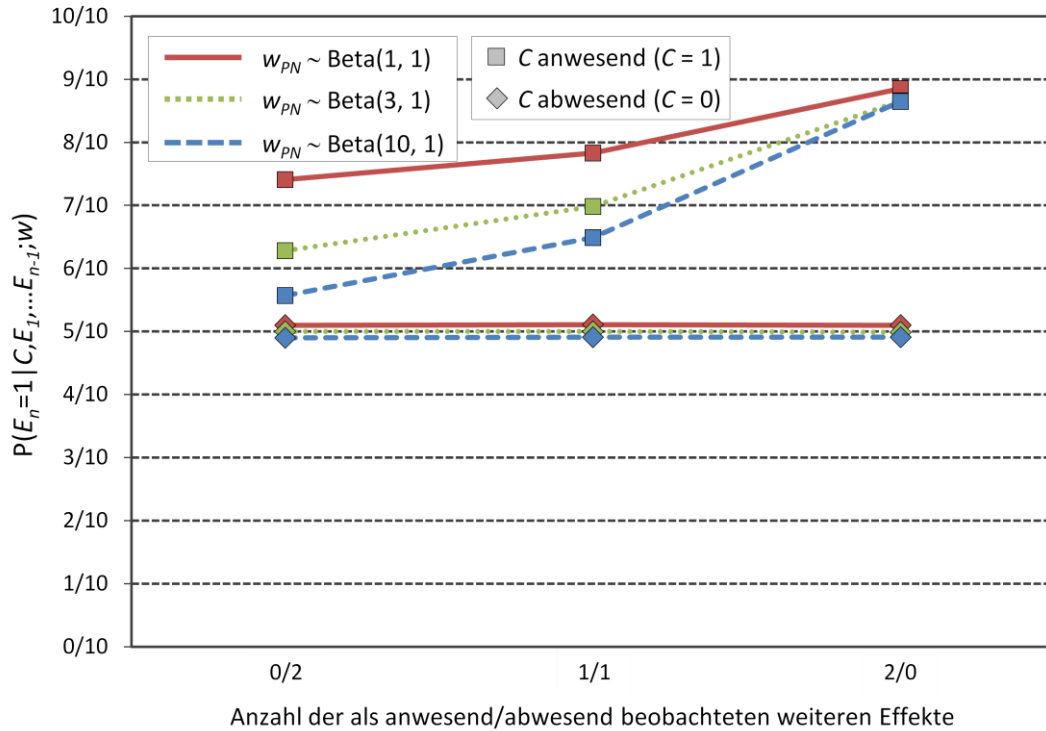


Abbildung 17: Modellvorhersage bei Annahme einer höheren Basisrate der Effekte (im Vergleich zu Abbildung 16) für jeweils unterschiedliche Stärken ( $w_{PN}$ ) der präventiven Ursache  $PN$  auf der Basis einer Monte-Carlo-Simulation mit 100 000 Durchgängen in einem *Common-Cause*-Modell mit einer Ursache  $C$  und drei Effekten. Die Parameter des kausalen System wurden hierfür aus folgenden Verteilungen gezogen:  $w_C \sim \text{Beta}(10, 1)$ ,  $b_E \sim \text{Beta}(1, 1)$  [statt  $b_E \sim \text{Beta}(1, 2)$ ] und  $b_{PN} \sim \text{Beta}(1, 1)$ ;  $b_C$  hat auf die Modellvorhersage keinen Einfluss. Die Vorhersagen für  $C = 0$  liegen übereinander; die drei Linien wurden daher mit einem Versatz (+0.1, 0, -0.1) versehen, um sie optisch sichtbar zu machen.

#### 4.4.2 Abhängigkeit von der Basisrate der Ursache ( $b_C$ )

Wie bereits im letzten Abschnitt dargelegt, ist die Modellvorhersage für die hier modellierte Zielinferenz unabhängig von der Basisrate der Ursache, da die Ursache immer beobachtet ist. Sie stellt daher im Hinblick auf die Modellierung von  $P(E_n = 1 | C, E_1, \dots, E_{n-1}; \omega)$  keinen relevanten freien Parameter dar.

#### 4.4.3 Abhängigkeit von der Kausalstärke von $C$ ( $w_C$ )

Die Modellvorhersage bei Abwesenheit von  $C$  ist – wie eben dargelegt – nur abhängig von  $b_E$  und damit unabhängig von  $w_C$ . Im Falle der Anwesenheit der gemeinsamen Ursache  $C$  gilt (siehe Tabelle 4):

$$b_E + w_C - w_C b_E \geq b_E + w_C(1 - w_{PN}) - w_C(1 - w_{PN})b_E \quad \forall w_{PN} \in [0, 1]. \quad (12)$$

Damit ist  $b_E + w_C - b_E w_C$  die Obergrenze der Modellvorhersage.

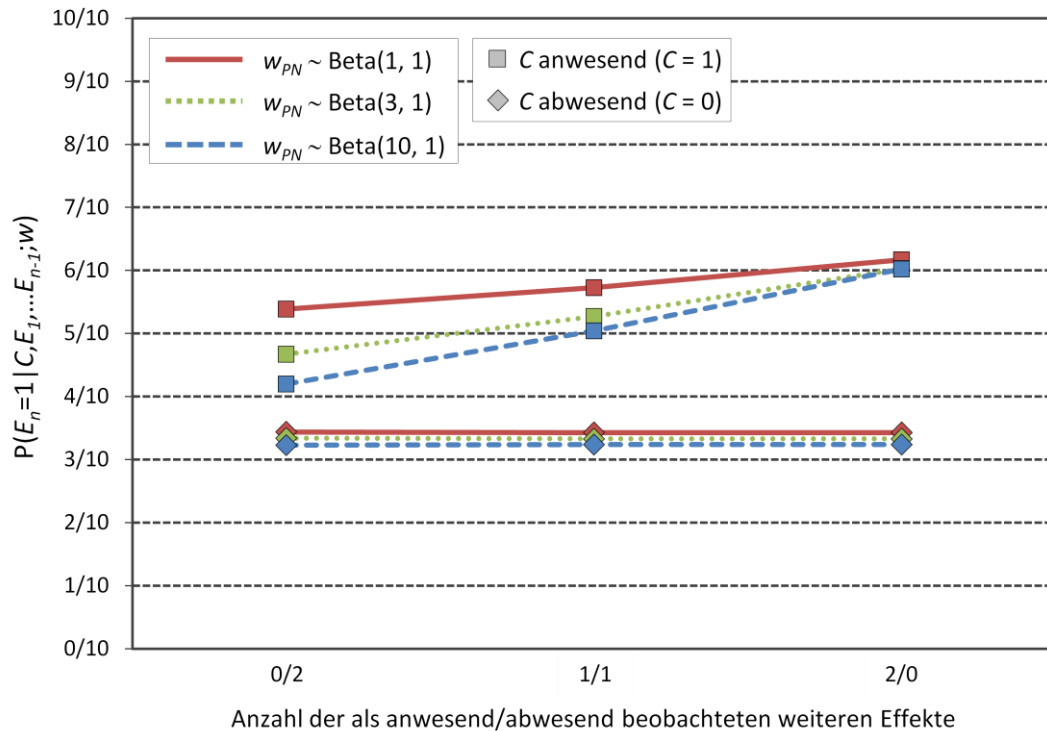


Abbildung 18: Modellvorhersage bei Annahme einer *niedrigeren Kausalstärke der gemeinsamen Ursache* (im Vergleich zu Abbildung 16) für jeweils unterschiedliche Stärken ( $w_{PN}$ ) der präventiven Ursache  $PN$  auf der Basis einer Monte-Carlo-Simulation mit 100 000 Durchgängen in einem *Common-Cause-Modell* mit einer Ursache  $C$  und drei Effekten. Die Parameter des kausalen Systems wurden hierfür aus folgenden Verteilungen gezogen:  $w_C \sim \text{Beta}(1,1)$  [statt  $w_C \sim \text{Beta}(10,1)$ ],  $b_E \sim \text{Beta}(1,2)$  und  $b_{PN} \sim \text{Beta}(1,1)$ ;  $b_C$  hat auf die Modellvorhersage keinen Einfluss.

Die Vorhersagen für  $C = 0$  liegen übereinander; die drei Linien wurden daher mit einem Versatz (+0.1, 0, -0.1) versehen, um sie optisch sichtbar zu machen.

Abbildung 18 zeigt die Vorhersage des Modells bei Annahme einer niedrigeren Stärke von  $C$ ,  $w_C$  wurde hierfür aus einer  $\text{Beta}(1,1)$ -Verteilung mit Erwartungswert 0.5 gezogen [statt aus einer  $\text{Beta}(10,1)$ -Verteilung mit einem Erwartungswert von 0.91]. Die Vorhersage ist deutlich gestaucht (weil durch die Absenkung von  $w_C$  die Range, in der sich die Vorhersage bewegen kann, von oben her kommend beschränkt wurde).

Eine genauere Analyse des Modells zeigt, dass der Parameter  $w_C$  zusammen mit dem Parameter  $b_E$  die Modellvorhersage auf den Bereich  $[b_E, b_E + w_C - b_E w_C]$  „skaliert“ und damit die qualitativ-strukturelle Modellvorhersage unbeeindruckt lässt. Der Parameter  $w_C$  beeinflusst zwar auch, wie stark die gemeinsame Fehlerursache  $PN$  als anwesend inferiert wird, nur beeinflusst dies die

Modellvorhersage außer in Extremfällen nur unwesentlich: Ist  $w_C$  z.B. sicher 1, unabhängige Fehler also sicher ausgeschlossen, so muss bei der Beobachtung bereits eines abwesenden Effekts und anwesender gemeinsamer Ursache die gemeinsame Fehlerquelle anwesend sein (als einzige Erklärungsmöglichkeit für die Abwesenheit des Effekts). Das bedeutet, dass sich die Vorhersage in diesen Fällen (mindestens ein Effekt abwesend) auf dem Niveau  $b_E + (1 - w_{PN}) - b_E(1 - w_{PN})$  bewegt und dann (alle Effekte anwesend) nach oben abknickt. (Einen ähnlichen Effekt findet man bei der Annahme einer extrem hohen Basisrate von  $PN$ ; siehe hierzu den kommenden Abschnitt und Abbildung 20.) Die Interaktion, z.B. zwischen dem *Sending*- und *Reading*-Fall, bleibt jedoch selbst in diesen Extremfällen erhalten. Die qualitativ-strukturelle Modellvorhersage ist folglich unabhängig von der Wahl von  $w_C$ .

#### 4.4.4 Abhängigkeit von der Basisrate der Fehlerursache ( $b_{PN}$ )

Die eigentliche Interaktion entsteht durch das Zusammenspiel der gemeinsamen Fehlerursache und den beobachteten Variablen. Im vorgestellten Modell werden die Annahmen über die Korrelation der kausalen Links im Parameter  $w_{PN}$  – der Stärke der gemeinsamen präventiven Ursache – abgebildet. Interessant scheint daher besonders, wie die Modellvorhersage auf eine Änderung der Annahmen bezüglich der Basisrate dieser Fehlerquelle,  $b_{PN}$ , reagiert. Für die Darstellung der Modellvorhersage in Abbildung 16 wurde dieser Parameter aus einer Beta(1,1)-Verteilung gezogen, also einer informativen *A-priori*-Verteilung mit einem Erwartungswert von 0.5.

Angenommen, die Anwesenheit der Fehlerquelle sei *a priori* sehr unwahrscheinlich (niedrige Basisrate), dann sollte das Modell insbesondere in weniger eindeutigen Fällen, z.B. ein Effekt anwesend und ein Effekt abwesend, seltener inferieren, dass die gemeinsame Fehlerquelle anwesend ist (sie ist ja *a priori* unwahrscheinlicher). Der Unterschied zwischen verschiedenen Manipulationen von  $w_{PN}$  sollte also sinken. In eindeutigen Fällen (z.B. beide Effekte abwesend) sollten die *A-priori*-Annahmen über die Anwesenheit von  $PN$  jedoch einen weniger starken Einfluss haben. Genau diesen Effekt sieht man deutlich in Abbildung 19. Zieht man die Basisrate aus einer Beta(10,1)-Verteilung mit einem Erwartungswert von 0.09, dann nimmt die Vorhersage eine eher konkave Form an. Die

grundsätzliche Interaktion bleibt jedoch deutlich erhalten. Die Interaktion würde nur in Extremfällen, hier insbesondere der sicheren Erwartung der Abwesenheit von  $PN$ , verschwinden, da dann  $PN$  keinen Einfluss mehr hätte und die Modellvorhersage der Vorhersage eines Markov-konformen Bayes-Netzes entspräche.

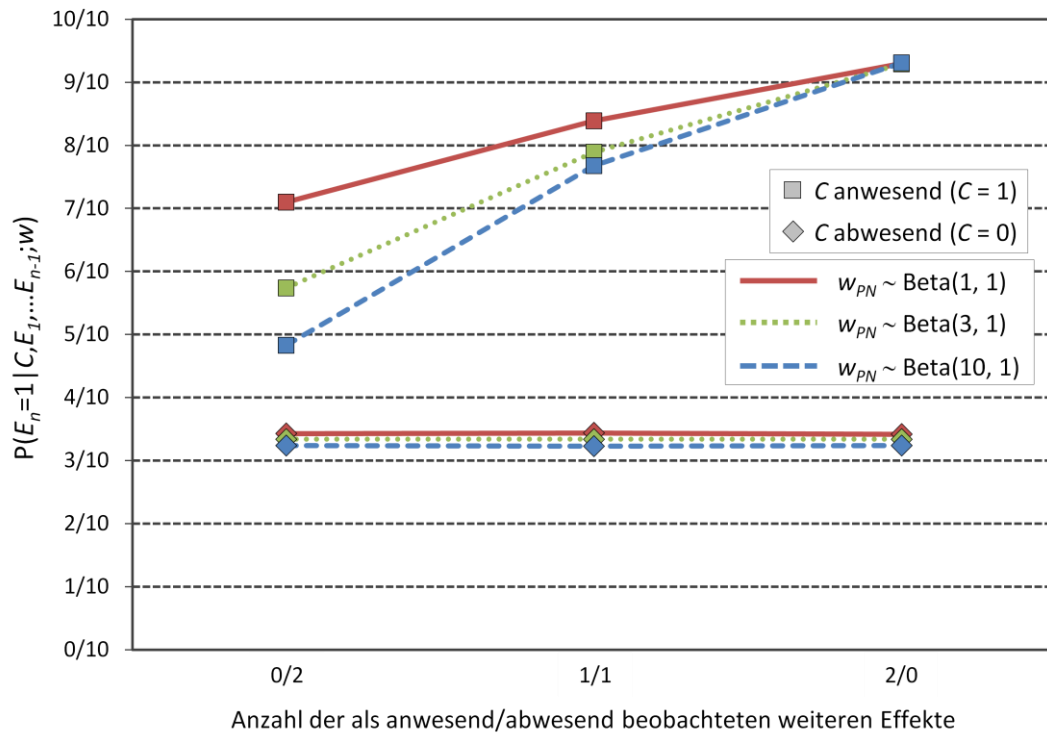


Abbildung 19: Modellvorhersage bei Annahme einer *sehr niedrigeren Basisrate* der präventiven Ursache  $PN$  (im Vergleich zu Abbildung 16) für jeweils unterschiedliche Stärken ( $w_{PN}$ ) der präventiven Ursache  $PN$  auf der Basis einer Monte-Carlo-Simulation mit 100 000 Durchgängen in einem *Common-Cause-Modell* mit einer Ursache  $C$  und drei Effekten. Die Parameter des kausalen Systems wurden hierfür aus folgenden Verteilungen gezogen:  $w_C \sim \text{Beta}(1,1)$ ,  $b_E \sim \text{Beta}(1,2)$  und  $b_{PN} \sim \text{Beta}(1,10)$  [statt  $b_{PN} \sim \text{Beta}(1,1)$ ];  $b_C$  hat auf die Modellvorhersage keinen Einfluss.

Die Vorhersagen für  $C = 0$  liegen übereinander; die drei Linien wurden daher mit einem Versatz (+0.1, 0, -0.1) versehen, um sie optisch sichtbar zu machen.

In Abbildung 20 ist kontrastierend die Modellvorhersage unter der *A-priori*-Annahme einer sehr hohen Basisrate der gemeinsamen Ursache dargestellt, indem  $b_{PN}$  aus einer  $\text{Beta}(10,1)$ -Verteilung mit einem Erwartungswert von 0.91 gezogen wurde. Die Anwesenheit der gemeinsamen Fehlerursache ist mithin also bereits *a priori* sehr wahrscheinlich. In weniger eindeutigen Fällen, ein Effekt anwesend, ein Effekt abwesend, sollte das Modell daher eher auf die Anwesenheit der gemeinsamen Fehlerquelle schließen; der Unterschied zwischen ver-

schiedenen  $w_{PN}$ -Manipulationen also steigen. Dies ist in Abbildung 20 auch gut zu erkennen. Insgesamt ist auch die Modellvorhersage aufgrund des abschwächenden Einflusses von  $PN$  nach unten hin gestaucht. Die Vorhersage einer Interaktion zwischen den  $w_{PN}$ -Manipulationen und der Anzahl der als anwesend beobachteten Effekte (bei Anwesenheit der gemeinsamen Ursache) bleibt jedoch bestehen. Die Vorhersage einer Interaktion würde nur im Extremfall der Annahme einer sicher immer anwesenden gemeinsamen Fehlerquelle verschwinden. Die Vorhersage entspräche dann der eines Bayes-Netzes mit unterschiedlichen Kausalstärken (drei horizontale Linien auf unterschiedlicher Höhe).

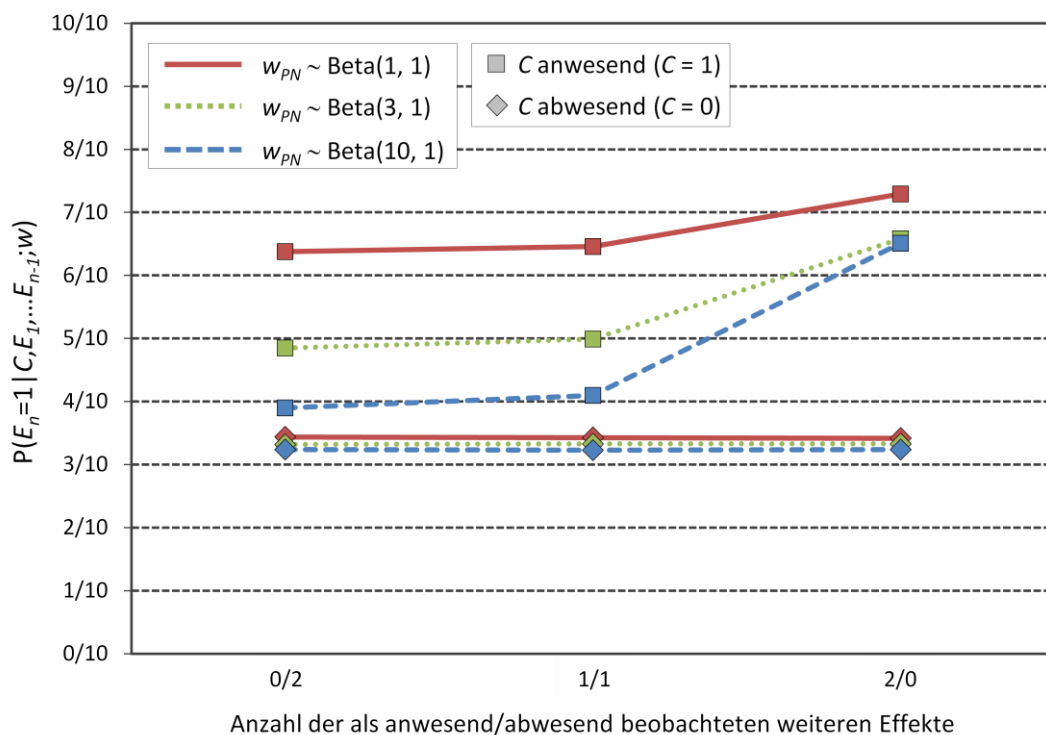


Abbildung 20: Modellvorhersage bei Annahme einer *sehr hohen Basisrate der präventiven Ursache PN* (im Vergleich zu Abbildung 16) für jeweils unterschiedliche Stärken ( $w_{PN}$ ) der präventiven Ursache  $PN$  auf der Basis einer Monte-Carlo-Simulation mit 100 000 Durchgängen in einem *Common-Cause-Modell* mit einer Ursache  $C$  und drei Effekten. Die Parameter des kausalen Systems wurden hierfür aus folgenden Verteilungen gezogen:  $w_C \sim \text{Beta}(1,1)$ ,  $b_E \sim \text{Beta}(1,2)$  und  $b_{PN} \sim \text{Beta}(10,1)$  [statt  $b_{PN} \sim \text{Beta}(1,1)$ ];  $b_C$  hat auf die Modellvorhersage keinen Einfluss.

Die Vorhersagen für  $C = 0$  liegen übereinander; die drei Linien wurden daher mit einem Versatz (+0.1, 0, -0.1) versehen, um sie optisch sichtbar zu machen.

Zusammenfassend kann also auch für die Basisrate der gemeinsamen Fehlerquelle,  $b_{PN}$ , konstatiert werden, dass sich die qualitativ-strukturelle Modellvorhersage von deren Wahl unbeeindruckt zeigt.

#### 4.5 Fitanalyse

Im vorhergehenden Abschnitt wurde qualitativ gezeigt, dass die strukturelle Modellvorhersage von den Annahmen über die Parameter relativ unabhängig ist. In diesem Abschnitt soll daher diese qualitative Analyse um eine quantitative ergänzt werden.

Das Modell hat drei „freie“ Parameter,  $w_C$ ,  $b_{PN}$  und  $b_E$ , die in der ursprünglichen Analyse händisch auf plausible Werte gesetzt worden sind. In Anlehnung an die Analysen von Roberts und Pashler (2000) ist die interessante Frage nun, in wie viel Prozent der Fälle man die mit den Daten konforme strukturelle Vorhersage des Modells erwarten darf, wenn die Parameter nicht an die Daten angepasst, sondern zufällig gezogen werden. Ist der resultierende Anteil groß, dann handelt es sich um eine robuste Vorhersage. Der Fit des Modells ist dann Ausdruck seiner guten Übereinstimmung mit dem zu erklärenden Phänomen und nicht schierer Ausdruck seiner Flexibilität (jeden beliebigen Datensatz zu fitten).

Betrachten wir für die Analyse nun die strukturelle Vorhersage an sich. Seien dafür  $R$  und  $S$  die beiden Bedingungen, die einer  $w_{PN}$ -Manipulation entsprechen, also z.B. „Reading“ vs. „Sending“.  $R_i^C$  bzw.  $S_i^C$  seien des Weiteren die Vorhersagen für Bedingung  $R$  bzw.  $S$  gegeben des Status der Ursache  $C$  (0 oder 1) und der Anzahl der weiteren als anwesend beobachteten Effekte  $i$  (0, 1 oder 2). Dann seien im Folgenden die strukturellen Vorhersagen:

1) *Haupteffekt von C*:  $\overline{RS}^{C=1} > \overline{RS}^{C=0}$ , d.h. im Durchschnitt sind die Vorhersagen für den Fall der Anwesenheit von Ursache  $C$  höher als im Falle von deren Abwesenheit (gerichteter Haupteffekt).

2) *Haupteffekt von i* gegeben  $C=1$ :  $\overline{RS}_0^{C=1} \leq \overline{RS}_1^{C=1} \leq \overline{RS}_2^{C=1}$  mit  $\exists i \in \{0,1\}: \overline{RS}_i^{C=1} < \overline{RS}_{i+1}^{C=1}$ , d.h. im Durchschnitt steigt im Falle der Anwesenheit die Vorhersage mit der Anzahl der als anwesend beobachteten Effekte monoton an (gerichteter Haupteffekt konditionalisiert auf  $C=1$ ).



3) Keine Effekte gegeben  $C=0$ :  $R_i^{C=0} = S_i^{C=0} \forall i$  sowie  $R_i^{C=0} = R_j^{C=0}, S_i^{C=0} = S_j^{C=0} \forall i, j$ , d.h. identische Vorhersagen für den Fall der Abwesenheit von  $C$  (keine Haupt- oder Interaktionseffekte konditionalisiert auf  $C=0$ )<sup>35</sup>.

4) Interaktion von  $R/S$  und  $i$  gegeben  $C=1$ :  $R_0^{C=1} - S_0^{C=1} \geq R_1^{C=1} - S_1^{C=1} \geq R_2^{C=1} - S_2^{C=1}$  mit  $\exists i \in \{0,1\}: R_i^{C=1} - S_i^{C=1} > R_{i+1}^{C=1} - S_{i+1}^{C=1}$ , d.h. gegeben der Anwesenheit von  $C$  nimmt der Unterschied zwischen den Bedingungen  $R/S$  mit zunehmender Anzahl der als anwesend beobachteten Effekte monoton ab (gerichteter Interaktionseffekt konditionalisiert auf  $C=1$ ). Zusammen mit Punkt 3 folgt aufgrund der Strenge dessen Formulierung auch die Vorhersage einer Drei-Wege-Interaktion.

Um nun zu prüfen, in wie vielen Fällen die Vorhersage des Modells bei zufälliger Ziehung der Parameter obigen Kriterien entspricht, wurde eine Simulation mit 10 000 Wiederholungen durchgeführt. Die drei Parameter  $w_C$ ,  $b_{PN}$  und  $b_E$  wurden dabei uniform aus ihrem Definitionsbereich (dem Intervall zwischen 0 und 1) gezogen und hierfür die Modellvorhersage bestimmt. Der Parameter  $w_{PN}$  wurde dabei auf 0.5 (Bedingung  $R$ : „Reading“) und 0.91 (Bedingung  $S$ : „Sending“) festgesetzt; dies entspricht dem Erwartungswert der beiden äußeren  $w_{PN}$ -Verteilungen, die in die Modellvorhersage in Abschnitt 4.3 eingegangen sind<sup>36</sup> und die die beiden experimentellen Bedingungen („Reading“ vs. „Sending“) am Besten repräsentieren.

In 99.2 % der Durchläufe entsprach die Modellvorhersage allen vier oben formulierten Bedingungen. 100 % der Fälle entsprachen Bedingung 1 (Haupteffekt von  $C$ ), 99.9 % der Fälle erfüllten Bedingung 2 (Haupteffekt von  $i$  gegeben  $C=1$ ), 100 % der Fälle entsprachen Bedingung 3 (keine Effekte für  $C=0$ ) und 99.2 % der Fälle erfüllten Bedingung 4 (Interaktionseffekt für  $C=1$ ).

Damit bestätigt die Simulationsstudie, dass die hohe qualitativ-strukturelle Übereinstimmung der Modellvorhersage mit den in Kapitel 3 präsentierten Daten nicht nur Ausdruck einer hohen Flexibilität des Modells ist. Die Vorhersage

<sup>35</sup> Diese Vorhersage könnte man auch weit weniger streng formulieren. Da das Modell aber mathematisch die strenge Version erfüllt und die gewählte Variante alle schwächeren Formulierungen umfasst, ist dies eine annehmbare Vereinfachung.

<sup>36</sup> Da die Modellvorhersage im vorliegenden Fall wie bereits dargelegt nur vom Erwartungswert abhängig ist, wurde in dieser Simulation auch nur mit diesem statt mit Verteilungen gearbeitet. Ansonsten wäre die Simulation nicht in zumutbarer Zeit durchzuführen gewesen. An den Ergebnissen ändert sich dadurch nichts.

des Modells ist in struktureller Hinsicht unabhängig von der Wahl der freien Parameter. Das Modell ist mithin inkompatibel mit allen anderen denkbaren Datensätzen und somit in höchstem Maße falsifizierbar. Die hohe Übereinstimmung zwischen Modellvorhersage und Daten ist also als „überzeugend“ im Sinne von Roberts und Pashler (2000) einzustufen.

#### 4.6 Zusammenfassung und Diskussion

Die in Kapitel 3 in den Basisexperimenten aufgezeigte Sensibilität der Versuchspersonen bezüglich zugrunde liegender kausaler Mechanismen wurde im vorliegenden Kapitel als flexible Fehlerattribution modelliert. Dazu wurde dem das kausale System repräsentierenden Bayes-Netz eine zusätzliche Fehlerquelle hinzugefügt, die gleichermaßen präventiv auf alle Links innerhalb einer *Common-Cause*-Struktur wirkt, und die Zielinferenz  $P(E_n = 1 | C, E_1, \dots, E_{n-1}; \omega)$  bayesianisch abgeleitet. Unterschiedliche Annahmen über die Stärke dieser gemeinsamen Fehlerquelle führen in diesem Modell zu unterschiedlich starken Abhängigkeiten der kausalen Links und damit zu unterschiedlich starken Markov-Verletzungen auf der Ebene der beobachteten Variablen. Die Zielinferenz kann dabei als zweistufiger Prozess gedacht werden. In einem ersten Schritt wird auf die Anwesenheit oder die Abwesenheit der gemeinsamen Fehlerursache geschlossen, je nachdem wie viele beobachtete Effekte der anwesenden Ursache abwesend sind – je mehr, desto unwahrscheinlicher, dass dies rein zufällig durch die unabhängigen Fehlerquellen zustande gekommen ist. In einem zweiten Schritt wird dann auf der Basis dieser Annahmen über die Anwesenheit der gemeinsamen Fehlerquelle auf die Wahrscheinlichkeit der Anwesenheit des Zieffekts geschlossen.

Die Modellvorhersage bildet dabei die Befunde sehr gut ab. Das Modell sagt für den Fall der Abwesenheit der gemeinsamen Ursache keinen Einfluss des Status der anderen Effekte und insbesondere keinen Einfluss verschiedener Annahmen über die zugrunde liegenden Kausalprozesse vorher, wie es sich auch in den Daten zeigt. Für den Fall der Anwesenheit der gemeinsamen Ursache sinkt die vorhergesagte Wahrscheinlichkeit der Anwesenheit des Zieffekts mit zunehmender Anzahl als abwesend beobachteter Effekte und zwar umso stärker je stärker die Annahme über die Abhängigkeit der kausalen Links ist. Dieser vorher-

---

gesagte bedingte Haupteffekt und die bedingte Interaktion finden sich ebenfalls sehr deutlich in allen drei Experimenten.

Die Ableitung des Modells wurde ergänzt durch eine Analyse der Modellflexibilität, also der Frage, inwieweit die gute Übereinstimmung der Modellvorhersage mit den Daten gegebenenfalls nur den freien Parametern und deren (wenn auch lediglich händischen) Anpassung an die Daten geschuldet ist. Dabei stellte sich heraus, dass die strukturelle Modellvorhersage (die relevanten Haupt- und Interaktionseffekte) sehr robust und *de facto* nicht von den freien Parametern abhängig ist. In einer Simulationsstudie mit 10 000 Durchgängen und zufälliger Ziehung der freien Parameter blieb diese in 99.2 % der Fälle stabil. Die gute Übereinstimmung des Modells mit den Daten ist damit im Sinne der Überlegungen von Roberts und Pashler (2000) trotz seiner freien Parameter überzeugend.

Die Einführung einer weiteren, mithin korrelierten Fehlerquelle in das kausale System scheint damit ein gutes Erklärungsmodell auf computationaler Ebene zu sein. Die Annahme, dass die Versuchspersonen kausale Systeme als Markov-konforme Bayes-Netzze repräsentieren, kann damit ggf. aufgegeben werden, ohne Bayes-Netzze und ihre positiven computationalen Eigenschaften als Teil der theoretischen Beschreibungssprache zu verlieren.

Nichtsdestotrotz steht ein expliziter Modelltest noch aus. Zwar wurde im vorliegenden Kapitel gezeigt, dass es eine gute Übereinstimmung zwischen Modell und Daten gibt und dass die gute Übereinstimmung nicht Ausdruck übermäßiger *Modellflexibilität* sein kann. Dies heißt aber nicht, dass der gute Fit nicht just Ausdruck einer hohen *Modellierungsflexibilität* im bayesianischen Framework ist (siehe dazu u.a. Waldmann et al., 2008). Es könnte schließlich sein, dass jedwedes empirische Ergebnis oder zumindest einige mögliche empirische Ergebnisse *post hoc* im Rahmen des bayesianischen Frameworks modelliert hätten werden können. In diesem Fall wäre eine gute Übereinstimmung zwischen Daten und Modell nicht sonderlich überraschend und damit auch nicht sonderlich überzeugend. In diesem Sinne sollen daher im folgenden Kapitel aus der grundsätzlichen Modellierungsidee der Einführung einer gemeinsamen Fehlerursache weitere Vorhersagen abgeleitet und getestet werden, um diesem Argument zu begegnen.

## 5 Modelltests

Wie in Abschnitt 4.6 beschrieben, steht ein expliziter Modelltest noch aus. Das Modell zeigt zwar eine gute Übereinstimmung mit den in den Basisexperimenten (Kapitel 3) gefundenen Befunden, die Entwicklung des Modells erfolgte jedoch auf der Basis eben der Ergebnisse, die das Modell erklären soll. Die Idee der Einführung einer gemeinsamen Fehlerquelle lässt sich aber gut auf andere Kontexte übertragen und die für diese gewonnenen Vorhersagen explizit empirisch testen.

So soll in Abschnitt 5.1 der Frage nachgegangen werden, inwieweit sich das Modell auf *Common-Cause*-Strukturen anwenden lässt, in denen beide Zustände der gemeinsamen Ursache kausal aktiv sind, und welche empirischen Vorhersagen sich daraus ergeben. In Abschnitt 5.2 wird die dem Modell zugrunde liegende Idee der Einführung einer gemeinsamen, an die Ursache gebundenen Fehlerquelle auf eine *Causal-Chain*-Struktur angewendet und damit überprüft, ob das Modell auch in anderen als *Common-Cause*-Strukturen, für die es entwickelt wurde, empirisch Bestand hat. In Abschnitt 5.3 soll schließlich getestet werden, welchen Einfluss es hat, wenn in eine *Common-Cause*-Struktur weitere beobachtbare präventive Ursachen eingeführt werden, die mit der modellendogenen Fehlerquelle bei der Fehlerattribution konkurrieren.

### 5.1 Schließen im Fall zweier kausal aktiver Zustände der Ursache

Die Experimente von Rehder und Burnett (2005, siehe dazu Abschnitt 2.2.3) haben eine Markov-Verletzung sowohl im Fall der Anwesenheit wie auch im Fall der Abwesenheit der gemeinsamen Ursache gezeigt. Dieser Befund scheint auf den ersten Blick im Widerspruch zu den Vorhersagen des in Kapitel 4 vorgestellten Modells zu stehen, sagt dieses schließlich konditionale Abhängigkeiten nur im Falle der Anwesenheit nicht jedoch im Falle der Abwesenheit der gemeinsamen Ursache vorher. Auch scheinen die Befunde von Rehder und Burnett (2005) sowie die in Kapitel 3 dargestellten Experimente sich in diesem Punkt zu widersprechen. Allerdings, wie bereits dargelegt, wird in beiden Paradigmen im unterschiedlichen Sinne vom kausalen Status der Variablen Gebrauch gemacht. Zwar sind alle Variablen binär, haben also nur zwei Zustände, und in beiden Paradigmen wird von anwesend und abwesend gesprochen, doch nur in den in Kapitel 3 vorgestellten Experimenten wird die Abwesenheit der Ursache klassisch als kau-

sal nicht aktiv beschrieben (in Korrespondenz zu der Idee von Cheng, 1997, dass ein abwesendes Ereignis nicht kausal wirksam im unmittelbaren Sinne sein kann). Rehder und Burnett (2005) verwendeten die Begriffe „abwesend“ und „anwesend“ jedoch im Hinblick auf den typischen Zustand eines Features, mithin war durch die Abwesenheit der typischen Ausprägung die Anwesenheit der ebenfalls kausal aktiven atypischen Ausprägung impliziert. Damit entfällt die Asymmetrie, die typischerweise in der Literatur zum kausalen Denken mit den Zuständen „abwesend“ und „anwesend“ der involvierten Ereignisse verbunden wird.

Im Folgenden sollen daher die beiden verschiedenen Interpretationen der binären Zustände einer kausal relevanten Variable als „0/1“-Fall (nur der Zustand 1 ist aktiv) und „A/B“-Fall (beide Zustände, A und B, sind kausal aktiv) bezeichnet werden. Das vorgestellte Modell behandelt derzeit nur den „0/1“-Fall, wie aus der bedingten Wahrscheinlichkeitstabelle (Tabelle 4 auf Seite 59) hervorgeht, und auch die Experimente beziehen sich nur auf den „0/1“-Fall. Im Folgenden sollen daher sowohl das Modell auf den „A/B“-Fall erweitert und die entsprechenden Vorhersagen abgeleitet, als diese auch im „*Mind-Reading-Alien*“-Paradigma getestet werden, indem dieses ebenfalls auf den „A/B“-Fall erweitert wird.

### 5.1.1 Modellierung

Das Modell, das im Abschnitt 4.2 für den „0/1“-Fall abgeleitet wurde, kann sehr einfach auf den „A/B“-Fall angewendet werden, indem die bedingte Wahrscheinlichkeitstabelle angepasst wird. Tabelle 4 beschreibt die bedingte Wahrscheinlichkeit der Anwesenheit eines Effekts gegeben der Anwesenheit bzw. Abwesenheit der gemeinsamen Ursache und gegeben der Anwesenheit bzw. Abwesenheit der gemeinsamen Fehlerquelle. Für den „A/B“-Fall muss nun definiert werden, was im A-Fall (Ursache im Zustand A) und was im B-Fall (Ursache im Zustand B) passiert. Sei dazu der A-Fall das Korrespondent zum 0-Fall und der B-Fall entsprechend zum 1-Fall. Dann seien also  $b_C$  und  $b_E$  die Basisrate der gemeinsamen Ursache  $C$  bzw. der Effekte  $E$ , mithin also die *A-priori*-Tendenz von  $C$  bzw.  $E$  im Zustand B zu sein. Dann sei  $w_C$  die Kausalstärke von  $C$  im Sinne der Wahrscheinlichkeit, mit der  $C$  seinen eigenen Zustand erfolgreich auf einen Effekt  $E$  überträgt

(und damit im Zweifel den *a priori* vorhandenen Zustand von  $E$  überschreibt)<sup>37</sup>. Mit dieser Interpretation ändert sich für die bedingten Wahrscheinlichkeiten  $P(E_n = B|C = B, PN)$  nichts, sie entsprechen den beiden bedingten Wahrscheinlichkeiten  $P(E_n = 1|C = 1, PN)$  im „0/1“-Fall (siehe Tabelle 4).

Für den Fall, dass die Ursache  $C$  sich im Zustand  $A$  befindet, ändert sich Folgendes: Die bedingte Wahrscheinlichkeit für  $E$  im Zustande  $B$  zu sein, wird durch zwei Dinge bestimmt: Zum einen durch die *A-priori*-Tendenz von  $E$  im Zustand  $B$  zu sein, was der Basisrate  $b_E$  entspricht, und zum anderen durch die Wahrscheinlichkeit, dass die Ursache  $C$  ihren Zustand *nicht* übertragen konnte (denn der ist ja  $A$ ), mithin also  $(1 - w_C)$  im Falle der Abwesenheit von  $PN$  bzw.  $(1 - w_C(1 - w_{PN}))$  im Falle der Anwesenheit von  $PN$ . Da es sich um zwei unabhängige Ereignisse handelt, werden die Wahrscheinlichkeiten multiplikativ verknüpft. Die vollständige Wahrscheinlichkeitstabelle findet sich in Tabelle 5.

Die eigentliche Modellvorhersage bestimmt sich aus Gleichung 6, wie sie in Abschnitt 4.2 für das Basismodell abgeleitet worden ist. In Abbildung 21 ist die Vorhersage dargestellt; die Parameter wurden dabei aus exakt den gleichen Verteilungen gezogen wie bei dem Basismodell im „0/1“-Fall. Gut zu erkennen ist, dass die Vorhersage für den Fall der Ursache im Zustand  $B$  exakt der Vorhersage im „0/1“-Fall für die Anwesenheit der Ursache entspricht. Der wesentliche Unterschied ergibt sich für den Fall der Ursache im Zustand  $A$ . Die bedingte Wahrscheinlichkeit des Zieffekts hängt nun ebenfalls vom Zustand der anderen Effekte ab und zwar umso stärker, je stärker die gemeinsame Fehlerquelle ist. Es ergibt sich mithin ein spiegelbildlich gestauchter Befund für die Ursachenzustände  $A$  und  $B$ , der über die *A-priori*-Tendenz der Effekte,  $A$  oder  $B$  zu sein, moduliert wird (im Fall einer Basisrate von 50% ergibt sich ein perfektes Spiegelbild an der „ $y=x$ “-Achse). Im Unterschied zum „0/1“-Fall sagt das Modell also auch für die Ursache im Zustand  $A$  einen Haupteffekt für die Anzahl der als im Zustand  $B$  beobachteten anderen Effekte sowie – vor allem – eine Interaktion zwischen diesem Faktor und der Stärke der gemeinsamen Fehlerursache („*Reading*“-vs.-„*Sending*“-Faktor) voraus.

<sup>37</sup> Zur Vereinfachung sei hier davon ausgegangen, dass der Zustand  $A$  der Ursache auch entsprechend den Zustand  $A$  beim Effekt verursacht. Das Mapping kann natürlich beliebig geändert werden, ohne dass sich an der Aussage etwas ändern würde.

Tabelle 5

Bedingte Wahrscheinlichkeit für Effekt  $E_i$ , im Zustand B zu sein, gegeben des Status der Ursache C und der gemeinsamen Fehlerquelle PN („A/B“-Fall)

C	PN	$P(E_i = B C, PN)$
A	0	$b_E(1 - w_C)$
A	1	$b_E(1 - w_C(1 - w_{PN}))$
B	0	$b_E + w_C - w_C b_E$
B	1	$b_E + w_C(1 - w_{PN}) - w_C(1 - w_{PN})b_E$

Anmerkungen.  $b_E$  – Basisrate des Effekts  $E_i$ , im Zustand B zu sein,  $w_C$  – Stärke der Ursache C,  $w_{PN}$  – Stärke der präventiven Ursache PN.

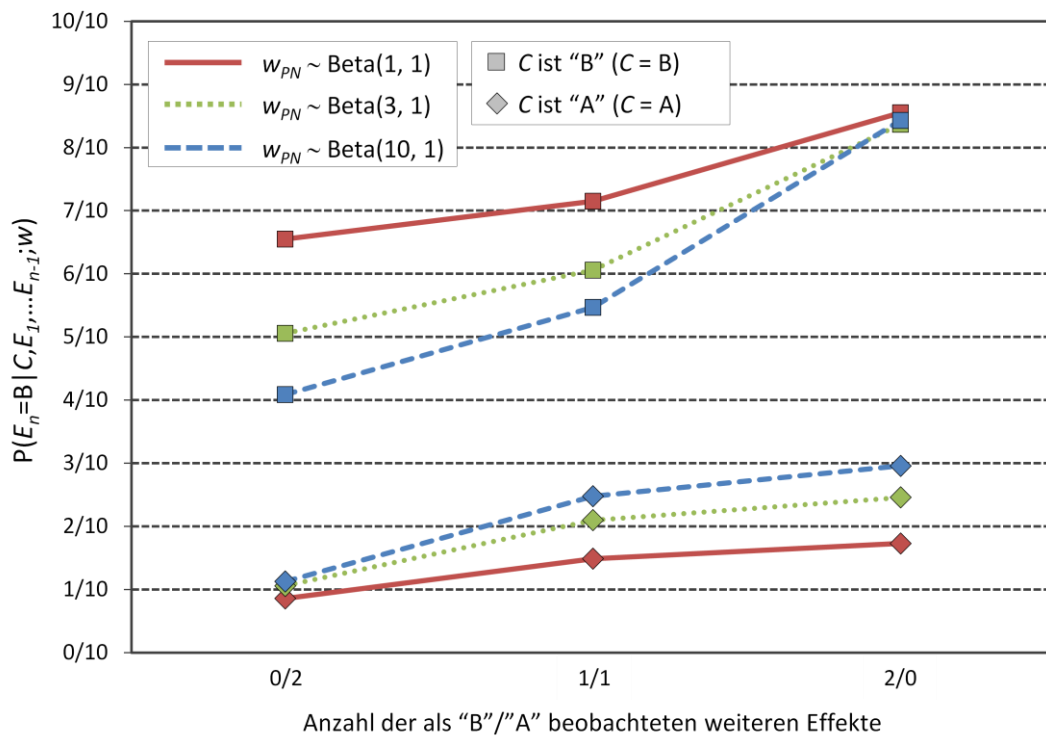


Abbildung 21. Modellvorhersage für unterschiedliche Stärken ( $w_{PN}$ ) der präventiven Ursache PN auf der Basis einer Monte-Carlo-Simulation mit 100 000 Durchgängen in einem *Common-Cause*-Modell mit einer Ursache C und drei Effekten im „A/B“-Fall. Die weiteren Parameter des kausalen Systems wurden hierfür aus folgenden Verteilungen gezogen:  $w_C \sim \text{Beta}(10, 1)$ ,  $b_E \sim \text{Beta}(1, 2)$  und  $b_{PN} \sim \text{Beta}(1, 1)$ ;  $b_C$  hat auf die Modellvorhersage keinen Einfluss.

### 5.1.2 Experiment 4: Überblick

Diese Vorhersage soll nun in einem Experiment getestet werden. Dafür kann einfach das in Kapitel 3 verwendete experimentelle Paradigma der Gedanken lesen- den bzw. Gedanken übertragenden Außerirdischen eingesetzt und angepasst

werden. Letztendlich handelt es sich beim „A/B“-Fall nur um eine andere Interpretation der kausalen Wirkung der Zustände der Variablen. Im „0/1“-Fall dachten die Aliens entweder an nichts oder an „POR“ und nur der Gedanke „POR“ konnte gesendet bzw. gelesen werden. Er entsprach damit dem kausal aktiven Zustand der Ursache („1“-Zustand). Die Instruktion kann insoweit aber leicht abgeändert werden, indem die Aliens an zwei distinkte Wörter, z.B. „TUS“ und „POR“, denken, die beide gesendet bzw. gelesen werden können. Diese einfache experimentelle Abänderung sollte zu einem dem Modell entsprechenden geänderten Antwortverhalten bei den Versuchspersonen führen.

### 5.1.3 Experiment 4: Methode

*Versuchspersonen.* An der Untersuchung nahmen 56 Personen teil (29 Frauen, 27 Männer; Durchschnittsalter 25.1 Jahre, Altersspanne 17–69 Jahre). Diese wurden auf dem Campus, im Zentralen Hörsaalgebäude (ZHG), im Psychologischen Institut, in den Mensen rund um den Campus der Universität Göttingen und zum Teil im Bekannten- und Verwandtenkreis der Versuchsleiterinnen und Versuchsleiter angeworben. Es handelte sich dabei größtenteils um Studierende der Universität Göttingen verschiedenster Fachrichtungen. Keiner der Probanden hatte bereits an einem der vorhergehenden Experimente der Serie teilgenommen. Die Teilnahme erfolgte freiwillig. Nach Ende der Untersuchung wurden die Teilnehmerinnen und Teilnehmer mit einer Süßigkeit belohnt und sofern gewünscht über den Zweck und das Ziel der Untersuchung genauer aufgeklärt.

Die Probanden wurden den beiden Bedingungen (Gedankenlesen vs. Gedankensenden) zufällig zugeteilt ( $N = 2 \times 28$ ).

*Versuchsdurchführung und Material.* Die Durchführung von Experiment 4 entsprach im Wesentlichen der Durchführung von Experiment 1. Das Experiment bestand wieder aus einer Instruktions- und einer Testphase. In der *Instruktionsphase* wurde den Probanden eine sechsseitige Instruktion vorgelegt, in welcher wieder erklärt wurde, dass es vier Aliens, Gonz, Murks, Brxxx und Zoohng gebe (das Bildmaterial entsprach dem von Experiment 1 und Experiment 2; siehe Abbildung 7). Anders als in den in Kapitel 3 vorgestellten Experimenten wurde den Versuchspersonen allerdings gesagt, dass die Aliens entweder an „TUS“ (Trinken



auf außerirdisch) oder an „POR“ (Essen auf außerirdisch) denken würden. Des Weiteren wurde erklärt, dass die Aliens viel häufiger an „TUS“ dächten als an „POR“. Damit entsprach der „TUS“-Gedanke von den beschriebenen Basisraten dem An-Nichts-Denken in den „0/1“-Experimenten.<sup>38</sup>

Überdies wurde dargelegt, dass Murks, Brxxx und Zoohng die Gedanken von Gonz lesen könnten („*Reading*“-Bedingung) bzw. Gonz seine Gedanken aus-senden und in die Köpfe von Murks, Brxxx und Zoohng übertragen könne („*Sending*“-Bedingung). Dies heiÙe im Konkreten, dass Murks, Brxxx und Zoohng häufig dann an „POR“ bzw. „TUS“ denken würden, wenn Gonz an „POR“ bzw. „TUS“ denke. Allerdings seien Murks, Brxxx und Zoohng nicht perfekt und manchmal unaufmerksam; und ein unaufmerksames Aliens schaffe es nicht, den Gedanken von Gonz zu lesen („*Reading*“-Bedingung) bzw. Gonz sei nicht perfekt, so dass er es nicht immer schaffe, seine Gedanken auszusenden und damit in die Köpfe von Murks, Brxxx und Zoohng zu übertragen („*Sending*“-Bedingung). In einem solchen Fall würden Murks, Brxxx und Zoohng jeder für sich genommen an „TUS“, manchmal aber auch an „POR“ denken, unabhängig davon, was Gonz gerade denke. Die Instruktion entsprach damit im Hinblick auf die mit den kausalen Prozessen verbundenen Fehlerstrukturen der deutlicheren Variante, wie sie in Experiment 1 und Experiment 2 verwendet wurde.

Daran anschließend wurden wieder am Paar Gonz und Murks die vier möglichen Fälle erklärt, die in der Erklärung von den vorhergehenden Experimenten abweichen, da nunmehr beide Gedanken von Gonz kausal aktiv sind: 1) *Gonz und Murks denken beide an „TUS“*. Zwar könne Murks auch aus freien Beweggründen an „TUS“ denken, viel wahrscheinlicher sei aber, dass Murks den Gedanken von Gonz gelesen habe („*Reading*“-Bedingung) bzw. Gonz seine Gedanken ausgesendet und damit in den Kopf von Murks übertragen habe („*Sending*“-Bedingung) und Murks darum an „TUS“ denke. 2) *Gonz und Murks denken beide an „POR“*. Zwar könne Murks auch aus freien Beweggründen an „POR“ denken, viel wahrscheinlicher sei aber, dass Murks den Gedanken von Gonz gelesen habe („*Rea-*

---

<sup>38</sup> Die Rolle des Denkens an „TUS“ (Trinken) und „POR“ (Essen) wurde über die Versuchspersonen und Bedingungen hinweg ausbalanciert. Für die Auswertung wurde zur besseren Vergleichbarkeit mit den Basisexperimenten alles derart rekodiert, dass „TUS“ von der Basisrate her „nichts“ und „POR“ entsprechend „POR“ entsprach.

*ding*“-Bedingung) bzw. Gonz seine Gedanken ausgesendet und damit in den Kopf von Murks übertragen habe („*Sending*“-Bedingung) und Murks darum an „POR“ denke. 3) *Gonz denkt an „POR“ und Murks denkt an „TUS“*. Gonz denke gerade an „POR“. Murks sei unaufmerksam und verpasse es sozusagen, den Gedanken von Gonz zu lesen („*Reading*“-Bedingung) bzw. Gonz sei aber nicht konzentriert genug und schaffe es nicht, seinen Gedanken auszusenden und damit in den Kopf von Murks zu übertragen („*Sending*“-Bedingung). In diesem Fall denke Murks also aus freien Beweggründen an „TUS“. 4) *Gonz denkt an „TUS“ und Murks denkt an „POR“*. Gonz denke gerade an „TUS“. Murks sei unaufmerksam und verpasse es sozusagen, den Gedanken von Gonz zu lesen („*Reading*“-Bedingung) bzw. Gonz sei aber nicht konzentriert genug und schaffe es nicht, seinen Gedanken auszusenden und damit in den Kopf von Murks zu übertragen („*Sending*“-Bedingung). In diesem Fall denke Murks also aus freien Beweggründen an „POR“, dies sei aber recht selten der Fall.<sup>39</sup>

Am Ende wurde verdeutlicht, dass die Beziehungen in gleicher Weise auch für Gonz und Brxxx sowie für Gonz und Zoohng gelten würden. Auf die Instruktion der Randwahrscheinlichkeiten, wie in Experiment 1, wurde verzichtet.

Vor Abschluss der Instruktionsphase sollten die Probanden die Instruktion in ihren wesentlichen Inhalten nochmal wiedergeben und diese dann bei Missverständnissen nochmals lesen.

In der *Testphase* wurden den Probanden wieder je sechs Testfälle in zufälliger Reihenfolge vorgelegt, in denen der Gedanke von Gonz und die Gedanken von zwei weiteren Aliens gezeigt wurden (diesmal jeweils entweder „TUS“ oder „POR“ anstatt „nichts“ oder „POR“; siehe für einen Beispiel-Testfall Abbildung 22). Aufgabe war es wieder, anzugeben, in wie vielen von zehn fiktiven Situationen mit der vorgelegten Gedankenkonfiguration das Zielalien an „POR“ denken werde (Skala von 0 bis 10). Die Testfälle waren wieder so konstruiert, dass je Zustand der gemeinsamen Ursache (also Gonz denkt an „TUS“ oder Gonz denkt an „POR“) drei Fälle abgedeckt wurden: Beide beobachteten Effektaliens denken an „TUS“, eins der Effektaliens denkt an „TUS“ und das andere an „POR“ sowie bei-

---

<sup>39</sup> Der letzte Halbsatz wurde angefügt, um die Asymmetrie in den Basisraten nochmals zu verdeutlichen.

de Effektliaens denken an „POR“. Der Gedanke welches Aliens unbeobachtet war, wurde zufällig bestimmt.

Für die Durchführung des Versuchs wurde ein Aktenordner angelegt, der die Instruktionen (je sechs Seiten für die „Reading“- bzw. „Sending“-Bedingung jeweils einmal für die „TUS“/„POR“- und die „POR“/„TUS“-Balancierung; also insgesamt vier Versionen), die 24 möglichen Testfälle sowie die Datentabelle und eine mit Excel erzeugte Randomisierungs- und Rekodierungstabelle enthielt (die 24 Testfälle wurden für beide Balancierungsbedingungen benutzt, nur mit jeweils unterschiedlicher Bedeutung). Die Rekodierung der Balancierungsbedingungen erfolgte beim Eintragen der Daten in die Datentabelle durch den Versuchsleiter anhand der Rekodierungstabelle; die Bedingungen wurden dergestalt rekodiert, dass von der Basisrateninstruktion her der „TUS“-Gedanke dem An-nichts-Denken in den Experimenten des „0/1“-Falles entsprach.

*Design.* Das Design entspricht einem 2x2x3-ANOVA-Design mit einem experimentellen Faktor („Reading“ vs. „Sending“; Faktor  $RS$ ) und zwei Messwiederholungsfaktoren (Status der Ursache  $C$ , sowie die Anzahl der als im Zustand „POR“ beobachteten weiteren Effekte, Faktor  $N_{E+}$ ).

*Vorhersage.* Die empirische Vorhersage ergibt sich aus der Modellvorhersage des an den „A/B“-Fall angepassten Modells (siehe Abschnitt 5.1.1). Es sollte sich daher wie in den Basisexperimenten für die Frage der Wahrscheinlichkeit, dass der Zieleffekt sich im Zustand B befindet, eine deutlich höhere Bewertung ergeben, wenn die Ursache sich im Zustand B befindet als wenn diese im Zustand A ist (Haupteffekt für Faktor  $C$ ). Des Weiteren wird wieder vorhergesagt, dass im Fall der Ursache im Zustand B (entspricht dem anwesendem Zustand in den vorherigen Experimenten) die Anzahl der weiteren im B-Zustand befindlichen Effekte einen deutlichen positiven Einfluss auf die Bewertungen haben sollte (Haupteffekt für den Faktor  $N_{E+}$  für  $C=B$ ) und dies in der „Sending“-Bedingung deutlich stärker als in der „Reading“-Bedingung (Interaktionseffekt  $RS \times N_{E+}$  für  $C=B$ ).

Im Unterschied zu den Basisexperimenten gilt diese Vorhersage nun aber auch für den Fall, dass die Ursache sich im A-Zustand befindet (entspricht dem abwesenden Zustand). Das heißt, es wird auch hier sowohl vorhergesagt, dass

die Bewertungen ansteigen, je mehr Effekte als im Zustand B befindlich beobachtet werden (Haupteffekt für den Faktor  $N_{E+}$  für  $C=A$ ) und dies in der „*Sending*“-Bedingung deutlich ausgeprägter als in der „*Reading*“-Bedingung ist (Interaktionseffekt  $RS \times N_{E+}$  für  $C=A$ ). Aufgrund der Symmetrie der Effekte wird eine deutlich ausgeprägte Drei-Wege-Interaktion nicht erwartet.



Abbildung 22. Ein Beispiel-Testfall, wie er in der Testphase von Experiment 4 benutzt wurde. Die Versuchspersonen hatten sich zehn Situationen mit der gegebenen Konfiguration vorzustellen und dann anzugeben, in wie vielen dieser Situationen das Zielalien (hier Brxxx) an „POR“ denken würde (Skala von 0 bis 10).

#### 5.1.4 Experiment 4: Ergebnisse

Die Ergebnisse von Experiment 4 sind in Abbildung 23 sowie in Tabelle 6 dargestellt; die detaillierten varianzanalytischen Ergebnisse finden sich in Tabelle E-1, Tabelle E-2 und Tabelle E-3 in Anhang E.

Die Bewertungen hinsichtlich der Wahrscheinlichkeit des unbeobachteten Zieleffekts, sich im Zustand B („POR“) zu befinden, sind im Durchschnitt deutlich höher, wenn sich die Ursache ebenfalls im Zustand B befindet (Haupteffekt C:  $F_{1,54}=146.05$ ,  $p<.001$ ,  $\eta^2=.73$ ).

Für den Fall, dass die Ursache sich im Zustand B befindet (Gonz denkt an „POR“; die oberen beiden Linien in Abbildung 23), hängen die Bewertungen deutlich von der Anzahl der weiteren sich im Zustand B befindlichen Effekte ab (Haupteffekt  $N_{E+}$  für C=B:  $F_{2,108}=31.47$ ,  $p<.001$ ,  $\eta^2=.37$ ). Ausgehend von null weiteren, sich im Zustand B befindenden Effekten steigt die Wahrscheinlichkeitseinschätzung mit deren zunehmender Anzahl in der „Reading“-Bedingung von 6.29 auf 7.64 ( $F_{2,54}=4.35$ ,  $p<.05$ ,  $\eta^2=.14$ ) und in der „Sending“-Bedingung von 3.54 auf 7.96 ( $F_{2,54}=38.00$ ,  $p<.001$ ,  $\eta^2=.58$ ). Der offensichtliche Anstieg ist mithin wie vorhergesagt in der „Sending“-Bedingung deutlich stärker ausgeprägt als in der „Reading“-Bedingung (Interaktionseffekt  $RS \times N_{E+}$  für C=B:  $F_{2,108}=8.94$ ,  $p<.001$ ,  $\eta^2=.14$ ).

Im Fall, dass die Ursache sich im Zustand A befindet (Gonz denkt an „TUS“; die beiden unteren Linien in Abbildung 23), hängen die Wahrscheinlichkeitseinschätzungen wie vorhergesagt ebenfalls deutlich von der Anzahl der im Zustand B befindlichen Effekte ab (Haupteffekt  $N_{E+}$  für C=A:  $F_{2,108}=20.25$ ,  $p<.001$ ,  $\eta^2=.27$ ). Dies gilt sowohl für die „Reading“-Bedingung, in der die Werte mit der Anzahl im Zustand B befindlicher Effekte von 1.93 auf 3.00 ansteigen ( $F_{2,54}=3.32$ ,  $p<.05$ ,  $\eta^2=.11$ ), als auch für die „Sending“-Bedingung, in der die Werte von 1.07 auf 3.82 steigen ( $F_{2,54}=19.91$ ,  $p<.001$ ,  $\eta^2=.42$ ). Damit ist die Abhängigkeit wie vorhergesagt und in Abweichung zu den Ergebnissen in den Basisexperimenten in der „Sending“-Bedingung deutlich stärker ausgeprägt als in der „Reading“-Bedingung (Interaktionseffekt  $RS \times N_{E+}$  für C=A:  $F_{2,108}=4.20$ ,  $p<.05$ ,  $\eta^2=.07$ ).

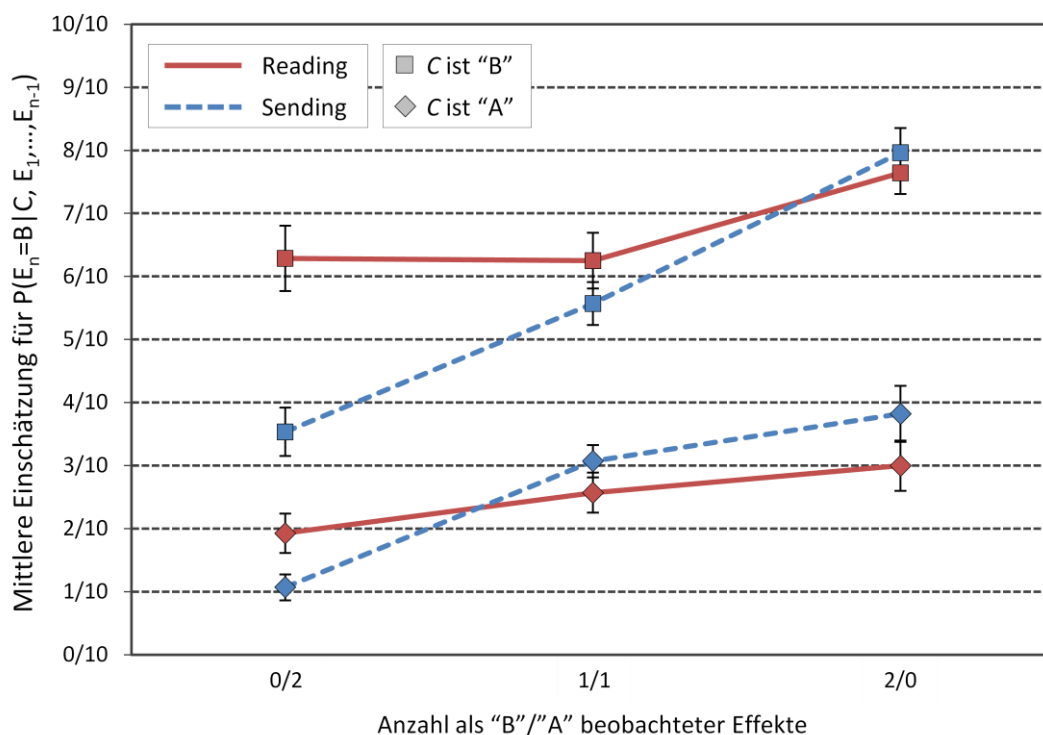
Eine Drei-Wege-Interaktion kann wie erwartet nicht nachgewiesen werden (Interaktionseffekt  $C \times RS \times N_{E+}$ :  $F_{2,108}=1.37$ ,  $p=.26$ ).

Tabelle 6

## Durchschnittliche Bewertung der Anwesenheit des Zieleffekts (Experiment 4)

	„Reading“-Bedingung (n = 28)			„Sending“-Bedingung (n = 28)		
	$N_{E^+} = 0$	$N_{E^+} = 1$	$N_{E^+} = 2$	$N_{E^+} = 0$	$N_{E^+} = 1$	$N_{E^+} = 2$
C = „B“ („POR“)	6.29 (2.73)	6.25 (2.34)	7.64 (1.79)	3.54 (2.03)	5.57 (1.79)	7.96 (2.06)
C = „A“ („TUS“)	1.93 (1.65)	2.57 (1.69)	3.00 (2.11)	1.07 (1.09)	3.07 (1.36)	3.82 (2.36)

*Anmerkungen.* Die Tabelle zeigt die durchschnittlichen Bewertungen auf einer Skala von „0 von 10“ bis „10 von 10“ (Häufigkeitsskala mit 11 Stufen) bezüglich der Wahrscheinlichkeit der Anwesenheit des Zieleffekts (Standardabweichung in Klammern). C = Zustand der gemeinsamen Ursache;  $N_{E^+}$  = Anzahl der als anwesend beobachteten weiteren Effekte.



*Abbildung 23.* Durchschnittliche Ratings bezüglich der Häufigkeit des Zieleffektzustandes B in zehn fiktiven Situationen gegeben dem Status der gemeinsamen Ursache C und dem Status der anderen beobachteten Effekte (jeweils intraindividuelle Variation) über den „Reading“-vs.-„Sending“-Faktor (interindividuelle Variation,  $N = 2 \times 28$ ) in Experiment 4 („A/B“-Fall). Die Fehlerbalken stellen den Standardfehler des Mittelwerts dar.

### 5.1.5 Diskussion

Die Ergebnisse bestätigen sehr deutlich die aus dem an den „A/B“-Fall angepassten Modell abgeleitete Vorhersage. Es konnte gezeigt werden, dass die Versuchspersonen sensitiv sind für eine Manipulation der kausalen Bedeutung der Zustände der Ursache. Der Unterschied zwischen der „Reading“- und der „Sending“-Bedingung fand sich nunmehr für beide Zustände der Ursache, und für beide Zustände zeigte sich eine deutliche Abhängigkeit der Bewertungen vom Status der weiteren Effekte. Der Befund repliziert damit im Prinzip auch die Ergebnisse von Rehder und Burnett (2005), die in ihren Experimenten ebenfalls symmetrische Markov-Verletzungen für beide Zustände der Ursache gefunden haben. Das in Kapitel 4 entwickelte Modell kann damit auch als Erklärungsmodell für die Rehder-und-Burnett-Befunde angesehen werden. In Kapitel 3 wurde in den Basisexperimenten zwar bereits das Erklärungsmodell von Rehder und Burnett widerlegt, da dieses die asymmetrischen Befunde nicht erklären kann. Noch blieb aber offen, wie die Befunde von Rehder und Burnett mit dem in Kapitel 4 abgeleiteten Modell in Einklang zu bringen sind, da dieses für die verschiedenen Zustände der Ursache asymmetrische Vorhersagen macht. Mit einer kleinen Anpassung der bedingten Wahrscheinlichkeitstabelle, die die Abhängigkeit des Effekts von den Zuständen der Ursache beschreibt, konnte das Modell aber auch auf den Fall zweier kausal aktiver Zustände der Ursache erweitert werden, ohne die ursprüngliche Modellierungsidee zu verändern. Experiment 4 kann damit als deutliche Bestätigung der Richtigkeit der dem Modell zugrunde liegenden Idee einer adaptiven Fehlerattribution gesehen werden.

## 5.2 Schließen in kausalen Ketten

Das in Kapitel 4 vorgestellte Modell wurde entwickelt, um kausale Schlussprozesse in *Common-Cause*-Strukturen abzubilden und besser auf computationaler Ebene zu beschreiben. Ein weiterer Nachweis der Richtigkeit des Modells, das Markov-Verletzungen als adaptive, auf den Annahmen über die zugrundeliegenden Kausalprozesse beruhende Fehlerattribution erklärt, könnte erbracht werden, wenn sich aus dem Modell auch testbare Hypothesen für andere kausale Strukturen ergäben. Das Modell enthält als strukturelle Komponente eine weitere Fehlerquelle, die die sich aus der Markov-Bedingung ergebenden unabhängigen

gen Fehlerterme für die Effekte einer gemeinsamen Ursache um eine gemeinsame Komponente erweitert. Der empirisch gefundene Unterschied zwischen einer „*Reading*“- und einer „*Sending*“-Bedingung (siehe Kapitel 3) wurde dabei auf unterschiedliche Annahmen in der Stärke dieser gemeinsamen Fehlerquelle zurückgeführt (siehe Kapitel 4). Je stärker die gemeinsame Fehlerquelle, desto stärker werden mögliche Fehler im kausalen System (z.B. anwesende Ursache, mehrere abwesende Effekte) auf die Ursache attribuiert und nicht unabhängig auf jeden einzelnen betroffenen Effekt (wie es die Markov-Bedingung erfordern würde). Eine Attribution auf die gemeinsame Fehlerquelle und damit auf die Seite der Ursache wirkt sich dann wiederum auf Schätzungen aus, die weitere, aber nicht beobachtete Effekte dieser Ursache betreffen. In einer solchen Situation würde man mithin starke Abhängigkeiten erwarten.

Da das Modell von der Idee her die gemeinsame Fehlerquelle an die Ursache bindet, lässt sich dies natürlich auch z.B. auf die Struktur einer kausalen Kette anwenden (siehe Abbildung 24 für ein Beispiel). Das letzte Element der Kette wäre der Zieleffekt. Nach der Markov-Bedingung dürfte das Urteil über den Zustand dieses Effekts nur von seiner direkten Ursache, nicht aber von seinen indirekten Ursachen abhängen. Führt man in dieses System die Idee einer gemeinsamen, an Ursachen gebundenen Fehlerquelle ein, so müsste jede Ursache in der Kette ihre eigene weitere Fehlerquelle bekommen. Da in einer echten kausalen Kette jede Ursache aber nur einen Effekt hat, wäre diese Fehlerquelle redundant. Weil es für die jeweiligen Ursache-Effekt-Paare nunmehr nicht mehr relevant ist, ob ein möglicher Fehler (Ursache anwesend, Effekt abwesend) auf die Seite der Ursache oder die Seite des Effekts attribuiert wird, dürfte sich die in *Common-Cause*-Strukturen gefundenen Dissoziation zwischen einer „*Reading*“- und einer „*Sending*“-Bedingung in kausalen Ketten nicht zeigen.

Ziel des vorliegenden Abschnitts ist daher, die Modellierungsidee auch formal auf eine kausale Kette anzuwenden, daraus Vorhersagen abzuleiten und in einem Experiment zu testen.

### 5.2.1 Modellierung

Wie eben bereits beschrieben kann die Basisidee der Einführung einer weiteren ursacheseitigen Fehlerquelle leicht auf die Struktur einer kausalen Kette ange-



wendet werden. Hierfür definieren wir für jeden Link  $X$ - $Y$  neben seiner Kausalstärke, die die unabhängige Fehlerquelle widerspiegelt, eine weitere Fehlerquelle  $PN_X$ , welche die korrelierte Fehlerquelle darstellt, die auf alle Effekte der Ursache  $X$  wirken würde, wenn diese mehrere Effekte hätte. Die Fehlerquellen  $PN_X$  sind wie im Basismodell unbeobachtet und mindern jeweils bei Anwesenheit die Effektivität der zugehörigen Ursache, ihre Effekte hervorzubringen. In Abbildung 15 ist eine solche um die Fehlerquellen  $PN_X$  erweiterte Kausale-Ketten-Struktur dargestellt. Die Variable  $E$  stellt hierbei das Ende der Kette, den ultimativen Effekt, dar,  $DC$  dessen direkte Ursache und die Variablen  $IC_1$  und  $IC_2$  die indirekten Ursachen von  $E$ . Die Parameter entsprechen den in Abschnitt 4.2 definierten Parametern; für die Verknüpfung von Ursache, Fehlerquelle und generativem Hintergrund gelte wieder die bedingte Wahrscheinlichkeitstabelle des Basismodells (Tabelle 4).

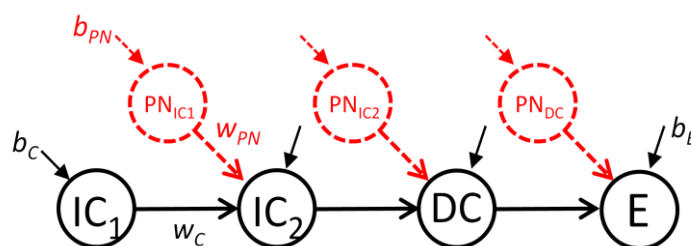


Abbildung 24: Ein für jede Ursache-Effekt-Relation  $X$ - $Y$  durch eine Fehlerursache  $PN_X$  erweitertes *Causal-Chain-Modell* mit vier Kettengliedern. Die Kausalstärke  $w_C$  gilt für alle Ursache-Effekt-Relationen, die Basisrate  $b_C$  gilt für alle Ursachen und die präventive Stärke  $w_{PN}$  gilt für alle Ursache-Effekt-Paare.

Ausgangspunkt der Modellierung ist wieder die Frage, die die Versuchspersonen in einem entsprechenden Experiment zu beantworten hätten, also die Wahrscheinlichkeit der Anwesenheit des nicht beobachteten Effekts  $E$  gegeben der anderen beobachteten Variablen des Systems (eine ausführlichere Variante der Herleitung findet sich in Anhang F)<sup>40</sup>:

$$P(E = 1 | DC, IC_1, IC_2; \omega) \quad (13)$$

Der Parametervektor  $\omega$  stellt dabei wieder das Vorwissen der Versuchspersonen über die kausale Domäne dar, welches die Annahmen über die Verteilun-

<sup>40</sup> Die Herleitung beschränkt sich auf den Spezialfall mit einer direkten und zwei indirekten Ursachen. Wie sich im Verlauf aber ergeben wird, gilt diese analog für beliebig viele indirekte Ursachen.

gen der kausalen Parameter beschreibt (Details siehe Abschnitt 4.2). Sie werden wieder in das Modell eingeführt und ausintegriert:

$$P(E = 1|DC, IC_1, IC_2; \omega) = \iint P(E = 1|DC, IC_1, IC_2; w) \cdot P(w|\omega) dw \quad (14)$$

mit

$$w = (w_C, w_{PN}, b_C, b_{PN}, b_E) \quad (15)$$

und analog unter der Annahme, dass<sup>41</sup>

$$P(w|DC, IC_1, IC_2; \omega) = P(w|\omega) \quad (16)$$

Nunmehr können die *PN*-Knoten eingeführt und, da sie unbeobachtet sind, durch Summation über die jeweils beiden Zustände  $PN_X = 0$  und  $PN_X = 1$  integriert und unter Anwendung der Markov-Bedingung vereinfacht werden<sup>42</sup>:

$$\begin{aligned} P(E_n = 1|DC, IC_1, IC_2; w) \\ = \sum_{PN_{DC}} \sum_{PN_{IC_1}} \sum_{PN_{IC_2}} P(E = 1|DC, PN_{DC}; w) \cdot P(PN_{DC}, PN_{IC_1}, PN_{IC_2}; w) \end{aligned} \quad (17)$$

Da die bedingte Wahrscheinlichkeit des Effekts nur vom Zustand der direkten Ursache und der dieser zugeordneten Fehlerquelle abhängt, nicht jedoch von den anderen Fehlerquellen, vereinfacht sich der Ausdruck weiter:

$$P(E_n = 1|DC, IC_1, IC_2; w) = \sum_{PN_{DC}} P(E = 1|DC, PN_{DC}; w) \cdot P(PN_{DC}; w) \quad (18)$$

Die Zielinferenz hängt damit nicht vom Status der indirekten Ursachen ab. Also gilt für Inferenzen im erweiterten Kettenmodell auf der Ebene der beobachteten Variablen weiterhin die Markov-Bedingung. Die erweiterte Struktur ist auf beobachteter Ebene einem Bayes-Netz äquivalent. In Anhang F wird gezeigt, dass dieses Bayes-Netz einer einfachen kausalen Kette ohne zusätzliche Fehlerquellen entspricht (siehe Abbildung 25). Deren Einfluss äußert sich in einem i.d.R. kleineren Kausalstärkeparameter, der sich wie folgt berechnet:  $w_C^* = w_C \cdot (1 - b_{PN}w_{PN})$ . Das heißt, die resultierende Kausalstärke ist umso kleiner, je *a priori* häufiger die Fehlerquelle *PN* ist und je stärker sie im Falle ihrer Anwesenheit wirkt. Für die Zielinferenz ergibt sich (siehe zum Vergleich Gleichung 9 auf Seite 61 für eine *Common-Cause*-Struktur mit Markov-Bedingung):

<sup>41</sup> Zur Begründung dieser Annahme siehe Abschnitt 4.2.

<sup>42</sup> Ausführlicher siehe Anhang F.

$$P(E_n = 1 | DC, IC_1, IC_2; w) = \begin{cases} b_E & | DC = 0 \\ w_C^* + b_E - w_C^* b_E & | DC = 1 \end{cases} \quad (19)$$

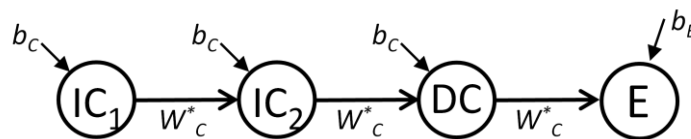


Abbildung 25. Eine einfache *Causal-Chain*-Struktur, die mit der um die ursachen- gebundenen Fehlerquelle erweiterten kausalen Kette identisch ist. Die Wirkung dieser weiteren, unbeobachteten Fehlerquellen äußert sich in einem kleineren Kausalstärkeparameter.

Damit stellt sich die Frage, welche Vorhersage im Falle einer Manipulation der Annahmen bezüglich der zugrundeliegenden Kausalprozesse erwartet werden kann. Eine Abhängigkeit von der Anzahl der als anwesend beobachteten indirekten Ursachen sagt das Modell offenkundig nicht vorher. Für den Fall der Abwesenheit der direkten Ursache sollte es keinen Unterschied zwischen den Bedingungen geben. Je nach Instruktion der Kausalstärke könnte es aber einen Unterschied im Falle der Anwesenheit der direkten Ursache geben, je nachdem inwieweit den beiden verschiedenen Kausalprozessen unterschiedliche Kausalstärken beigemessen werden. Sofern auf beobachteter Ebene, also  $w_C^*$  kein Unterschied instruiert wird<sup>43</sup>, sollte sich auch im Fall der Anwesenheit der Ursache kein Unterschied zwischen den Bedingungen zeigen.

### 5.2.2 Experiment 5: Überblick

Die Vorhersagen sollen nun in einem weiteren Experiment überprüft werden. Das in den vorhergehenden Experimenten verwendete Paradigma lässt sich aufgrund seiner Flexibilität wieder auf die vorliegende Situation anpassen. Ziel ist es, die Inferenzen über einen Zieleffekt in einer kausalen Kette in Abhängigkeit von seiner direkten Ursache und seinen indirekten Ursachen zu erfassen und dabei die Annahmen der Versuchspersonen über die zugrundeliegenden Kausalprozesse zu manipulieren. Letzteres kann wieder über die bekannte „*Reading*“-vs.-

<sup>43</sup> Diese Interpretation weicht etwas vom Basismodell ab. In diesem wurde  $w_C$  als über die experimentellen Bedingungen hinweg konstant angenommen; es änderte sich lediglich die Stärke der gemeinsamen präventiven Ursache  $PN$ . Diese Interpretation hätte eine unterschiedliche resultierende Kausalstärke  $w_C^*$  zur Folge. Sinnvoller erscheint es aber, im vorliegenden Fall von einer konstanten Kausalstärke  $w_C^*$  auszugehen. Auch im Basismodell ergeben sich aus dieser Annahme keine Änderungen der strukturellen Modellvorhersage.

„*Sending*“-Manipulation erreicht werden. Da im verwendeten Paradigma frei manipulierbar ist, wer von wem die Gedanken liest („*Reading*“-Bedingung) bzw. wer an wen seine Gedanken sendet („*Sending*“-Bedingung), lässt sich entsprechend einfach statt einer *Common-Cause*-Struktur eine *Causal-Chain*-Struktur instruieren, indem z.B. ein Alien jeweils die Gedanken seines linken Nachbarn liest („*Reading*“-Bedingung) bzw. seine Gedanken an seinen rechten Nachbarn aussendet („*Sending*“-Bedingung). Mit dieser Änderung der Instruktion sollten die in den Basisexperimenten gefundenen Unterschiede (v.a. die Interaktionen) im Hinblick auf die „*Reading*“-vs.-„*Sending*“-Manipulation und die Anzahl der beobachteten indirekten Ursachen verschwinden.

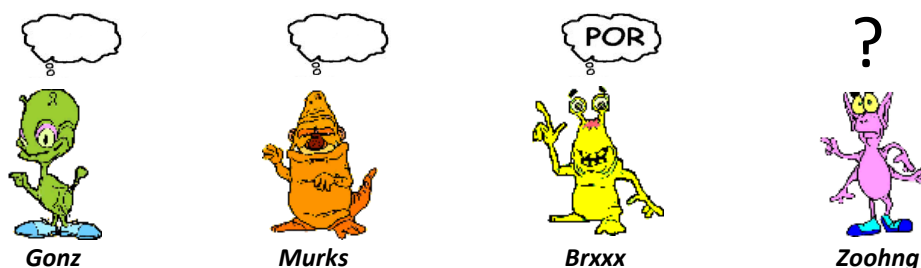
### 5.2.3 Experiment 5: Methode

*Versuchspersonen.* 50 Personen (22 Frauen, 28 Männer; Durchschnittsalter 24.3 Jahre, Altersspanne 19–40 Jahre) nahmen an der Untersuchung teil. Diese wurden auf dem Campus, im Zentralen Hörsaalgebäude (ZHG), im Psychologischen Institut, in den Mensen rund um den Campus der Universität Göttingen und zum Teil im Bekannten- und Verwandtenkreis der Versuchsleiterinnen und Versuchsleiter angeworben. Es handelte sich dabei größtenteils um Studierende der Universität Göttingen verschiedenster Fachrichtungen. Keiner der Probanden hatte bereits an einem der vorhergehenden Experimente der Serie teilgenommen. Die Teilnahme erfolgte freiwillig. Nach Ende der Untersuchung wurden die Teilnehmerinnen und Teilnehmer mit einer Süßigkeit belohnt und sofern gewünscht über den Zweck und das Ziel der Untersuchung genauer aufgeklärt.

Die Probanden wurden den beiden Bedingungen (Gedankenlesen vs. Gedankensenden) zufällig zugeteilt ( $N = 2 \times 25$ ).

*Versuchsdurchführung und Material.* Die Durchführung von Experiment 5 entsprach vom Grundschema her den vorhergehenden Experimenten, d.h. das Experiment bestand aus einer Instruktions- und einer Testphase. In der *Instruktionsphase* wurde den Probanden eine vierseitige Instruktion vorgelegt, in der erklärt wurde, dass es vier Aliens, Gonz, Murks, Brxxx und Zoohng, gebe, die meistens an nichts und manchmal an „POR“ (Essen auf außerirdisch) denken würden. Das Bildmaterial wurde von Experiment 3 übernommen (siehe Abbildung 12). Des

Weiteren wurde erklärt, dass sich gezeigt habe, alle Aliens hätten die Fähigkeit, die Gedanken des links von ihnen stehenden Aliens zu lesen („*Reading*“-Bedingung) bzw. ihre Gedanken an das rechts von ihnen stehende Alien zu senden und damit in dessen Kopf zu übertragen („*Sending*“-Bedingung). Im Konkreten heie dies, dass ein Alien hufig dann an „POR“ denke, wenn das Alien links von ihm an „POR“ denke. Allerdings seien die Aliens nicht perfekt. Manchmal seien sie unaufmerksam und ein unaufmerksames Alien schaffe es nicht, den Gedanken des links stehenden Aliens zu lesen („*Reading*“-Bedingung) bzw. seinen Gedanken auszusenden und damit in den Kopf des rechts stehenden Aliens zu bertragen („*Sending*“-Bedingung). In einem solchen Fall denke das Gedanken lesende Alien („*Reading*“-Bedingung) bzw. das Zielalien („*Sending*“-Bedingung) fr sich genommen manchmal an „POR“, unabhngig davon, was das Bezugsalien gerade denke.



*Abbildung 26.* Ein Beispiel-Testfall, wie er in der Testphase von Experiment 5 Verwendung fand. Die Versuchspersonen hatten sich zehn Situationen mit der gegebenen Konfiguration vorzustellen und dann anzugeben, in wie vielen dieser Situationen das Zielalien, Zoohng, an „POR“ denken wrde (Skala 0 bis 10). Im Unterschied zu den vorhergehenden Experimenten war das zugrundeliegende kausale System als Kette instruiert, in der das jeweils links stehende Alien die Gedanken des jeweils rechts stehenden Aliens verursacht.

Am Paar Gonz und Murks wurden wieder die vier mglichen Flle erklrt, wie dies in Experiment 1 und Experiment 2 gemacht wurde. Am Ende der Instruktionsphase hatten die Versuchspersonen wieder die wesentlichen Inhalte der Instruktion wiederzugeben und diese ggf. nochmals zu lesen.

In der Testphase wurden den Versuchspersonen acht Testflle in zuflliger Reihenfolge vorgelegt (fr einen Beispiel-Testfall siehe *Abbildung 26*). Das letzte Alien in der Kette, Zoohng, diente immer als Zielalien. Manipuliert wurden die Gedanken der direkten Ursache, Brxxx, sowie die Gedanken der beiden indirek-

ten Ursachen, Gonz und Murks. Insgesamt ergeben sich dadurch acht Fälle, die diesmal auch alle vorgelegt wurden.

Für die Durchführung des Versuchs wurde wieder ein Aktenordner angelegt, der die Instruktionen (je vier Seiten für die „*Reading*“- bzw. die „*Sending*“-Bedingung), die acht Testfälle sowie die Datentabelle und eine mit Excel erzeugte Randomisierungstabelle (für die randomisierte Reihenfolge) enthielt. Die Durchführung erfolgte auf dem Campus der Universität Göttingen und in umliegenden Gebäuden. Sie dauerte ca. 10–15 Minuten pro Versuchsperson.

*Design.* Für die Auswertung wurden die Bewertungen für die beiden Fälle, in denen eine der beiden indirekten Ursachen anwesend ist (entweder Gonz oder Murks denkt an „POR“), pro Versuchsperson per Durchschnittsbildung zusammengefasst. Dann ergibt sich wie in den vorhergehenden Experimenten ein 2x2x3-ANOVA-Design mit dem experimentellen Faktor *RS*, „*Reading*“ vs. „*Sending*“ sowie zwei Messwiederholungsfaktoren: dem Status der direkten Ursache (*direct cause, DC*) und der Anzahl der als anwesend beobachteten indirekten Ursachen (*indirect cause, N<sub>IC+</sub>*).

*Vorhersage.* Die empirische Vorhersage ergibt sich aus der Vorhersage des auf die *Causal-Chain*-Struktur angewendeten Basismodells. Wenn die direkte Ursache anwesend ist, sollten sich demnach deutlich höhere Bewertungen der Anwesenheit des Zieffekts zeigen, als wenn die direkte Ursache abwesend ist (Haupteffekt *DC*). Sowohl für den Fall der Anwesenheit als auch für den Fall der Abwesenheit der Ursache sollte sich kein Einfluss der Anzahl der als anwesend beobachteten indirekten Ursachen zeigen (kein Haupteffekt für den Faktor *N<sub>IC+</sub>*)<sup>44</sup>. Insbesondere sollte es keinen Unterschied in der Wirkung der Anzahl der als anwesend beobachteten indirekten Ursachen im Hinblick auf den „*Reading*“-vs.-„*Sending*“-Faktor (Faktor *RS*) geben – auch dann, wenn entgegen der Vorhersage doch ein Haupteffekt auftreten sollte. Ebenso wird eine Drei-Wege-Interaktion nicht erwartet.

---

<sup>44</sup> Wie es sich aber bereits in den vorhergehenden Experimenten gezeigt hat, mag ein solcher Haupteffekt auftreten, weil die Versuchspersonen eine weitere, generative Ursache annehmen. Eine solche müsste dann aber unabhängig von der „*reading*“-vs.-„*sending*“-Manipulation sein, d.h. die vorhergesagten Interaktionseffekte sind die bedeutenderen Effekte.

#### 5.2.4 Experiment 5: Ergebnisse

In Tabelle 7 und in Abbildung 27 sind die Ergebnisse von Experiment 5 dargestellt. Die detaillierten varianzanalytischen Ergebnisse finden sich in Tabelle G-1, Tabelle G-2 und Tabelle G-3 in Anhang G. Im Wesentlichen konnten die Vorhersagen gut bestätigt werden, insbesondere zeigt sich offenkundig keinerlei Einfluss der „Reading“-vs.-„Sending“-Manipulation.

Wie in den vorhergehenden Experimenten sind die Bewertungen hinsichtlich der Anwesenheit des Zieleffekts im Durchschnitt deutlich höher, wenn die direkte Ursache anwesend ist (Haupteffekt  $DC$ :  $F_{1,48}=191.99$ ,  $p<.001$ ,  $\eta^2=.80$ ).

Im Falle der Anwesenheit der Ursache (Brxxx denkt an „POR“; die oberen beiden Linien in Abbildung 27) steigen die Bewertungen ausgehend von null als anwesend beobachteter indirekter Ursachen insgesamt leicht, aber dennoch signifikant an (Haupteffekt  $N_{IC+}$  für  $DC=1$ :  $F_{2,96}=11.77$ ,  $p<.001$ ,  $\eta^2=.20$ ). In der „Reading“-Bedingung wachsen sie im Durchschnitt schwach monoton von 6.64 auf 7.16 ( $F_{2,48}=8.66$ ,  $p<.01$ ,  $\eta^2=.27$ ). In der „Sending“-Bedingung steigen sie von 6.48 auf 7.36 ( $F_{2,48}=5.56$ ,  $p<.01$ ,  $\eta^2=.19$ ). Ein Unterschied in der Abhängigkeit im Hinblick auf die experimentelle Manipulation findet sich nicht (Interaktionseffekt  $RS \times N_{IC+}$  für  $DC=1$ :  $F_{2,96}<1$ ,  $p=.50$ ), wie vorhergesagt.

Im Falle der Abwesenheit der Ursache (Brxxx denkt an nichts; die unteren beiden Linien in Abbildung 27) steigen die Bewertungen ebenfalls leicht, in zum Fall der Anwesenheit der Ursache vergleichbarem Umfang, ebenso signifikant an (Haupteffekt  $N_{IC+}$  für  $DC=0$ :  $F_{2,96}=6.47$ ,  $p<.01$ ,  $\eta^2=.12$ ); in der „Reading“-Bedingung von 1.88 auf 2.36 ( $F_{2,48}=2.61$ ,  $p=.08$ ,  $\eta^2=.10$ ) und in der „Sending“-Bedingung von 2.00 auf 2.80 ( $F_{2,48}=3.88$ ,  $p<.05$ ,  $\eta^2=.14$ ). Eine Interaktion zwischen dem „Reading“-vs.-„Sending“-Faktor und der Anzahl als anwesend beobachteter indirekter Ursachen (Faktor  $N_{IC+}$ ) findet sich nicht (Interaktionseffekt  $RS \times N_{IC+}$  für  $DC=0$ :  $F_{2,96}<1$ ,  $p=.66$ ).

Eine Drei-Wege-Interaktion kann wie erwartet ebenfalls nicht nachgewiesen werden (Interaktionseffekt  $DC \times RS \times N_{IC+}$ :  $F_{2,96}<1$ ,  $p=.99$ ).

Tabelle 7

## Durchschnittliche Bewertung der Anwesenheit des Zieleffekts (Experiment 5)

	„Reading“-Bedingung (n = 25)			„Sending“-Bedingung (n = 25)		
	$N_{IC^+} = 0$	$N_{IC^+} = 1$	$N_{IC^+} = 2$	$N_{IC^+} = 0$	$N_{IC^+} = 1$	$N_{IC^+} = 2$
$DC = 1$ („POR“)	6.64 (2.08)	6.64 (1.95)	7.16 (1.97)	6.48 (2.29)	6.72 (1.87)	7.36 (1.87)
$DC = 0$ (nichts)	1.88 (1.39)	2.12 (0.94)	2.36 (1.32)	2.00 (1.04)	2.44 (1.11)	2.80 (1.87)

Anmerkungen. Die Tabelle zeigt die durchschnittlichen Bewertungen auf einer Skala von „0 von 10“ bis „10 von 10“ (Häufigkeitsskala mit 11 Stufen) bezüglich der Wahrscheinlichkeit der Anwesenheit des Zieleffekts (Standardabweichung in Klammern).  $DC$  = Zustand der direkten Ursache;  $N_{IC^+}$  = Anzahl der als anwesend beobachteten indirekten Ursachen.

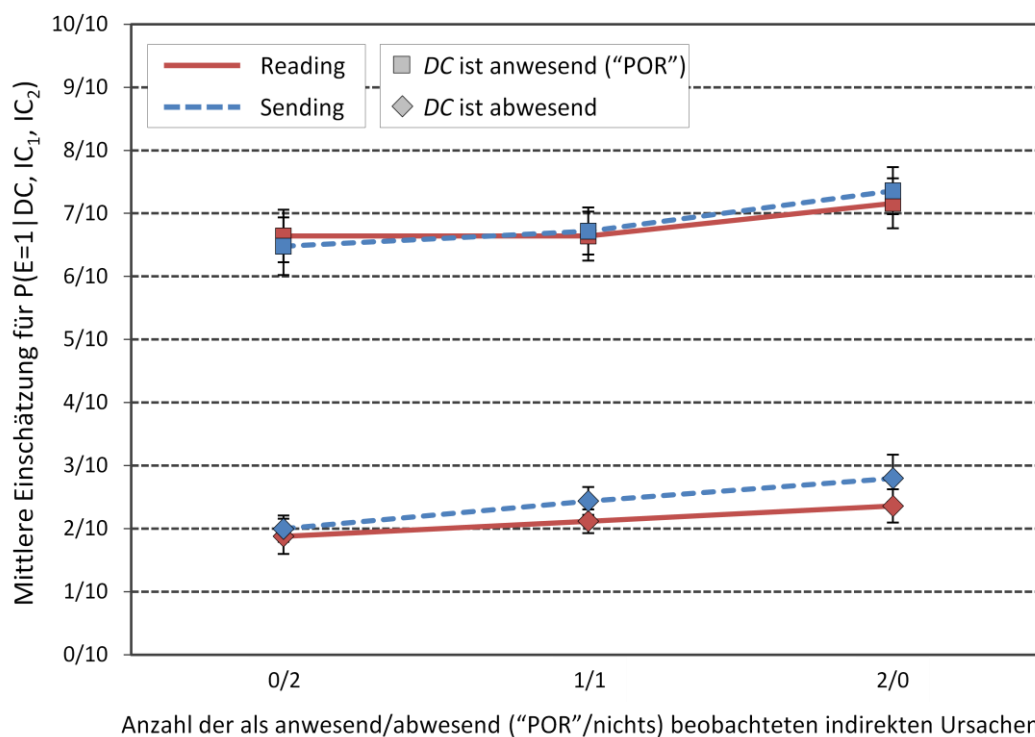


Abbildung 27. Durchschnittliche Ratings bezüglich der Häufigkeit der Anwesenheit des Zieleffekts in zehn fiktiven Situationen gegeben dem Status der direkten Ursache  $DC$  und dem Status der anderen beobachteten Effekte (jeweils intraindividuelle Variation) über den „Reading“-vs.-„Sending“-Faktor (interindividuelle Manipulation,  $N = 2 \times 25$ ) in Experiment 5. Die Fehlerbalken stellen den Standardfehler des Mittelwerts dar.



### 5.2.5 Diskussion

Die Ergebnisse von Experiment 5 bestätigen sehr deutlich die Vorhersage des auf die *Causal-Chain*-Struktur angepassten Modells: Ein Unterschied zwischen den beiden experimentellen Bedingungen („*Reading*“ vs. „*Sending*“) konnte weder im Falle der Anwesenheit noch im Falle der Abwesenheit der gemeinsamen Ursache gefunden werden. Damit zeigt sich sehr deutlich, dass die Inferenzen der Versuchspersonen wirklich von der Kausalstruktur und deren Eigenschaften abhängig sind. Damit ist auch belegt, dass die eingeführte zusätzliche Fehlerquelle mit der jeweiligen Ursache assoziiert, also an den dort ausgelösten Kausalprozess gekoppelt ist. Die vorgeschlagene erweiterte Fehlerstruktur ist damit intrinsische Eigenschaft der Ursache-Effekt-Relation und nicht lediglich (irgend-)eine unbeobachtete korrelierte Fehlerquelle, die an das kausale System angedockt wurde (und dort auf der Ebene der beobachteten Variablen Markov-Verletzungen generiert).

Nichtsdestotrotz sind auch in Experiment 5 wieder Abhängigkeiten von der Anzahl der als anwesend beobachteten weiteren Variablen (hier: indirekte Ursachen) aufgetreten, die wieder unabhängig von der Manipulation bezüglich der zugrunde liegenden Mechanismen sind. Wie bereits in einigen vorhergehenden Experimenten, die in Kapitel 3 beschrieben wurden, scheinen die Versuchspersonen wieder eine weitere, unbeobachtete generative Ursache zu inferieren, die auf alle Variablen wirkt (wie z.B. die Vermutung, dass deshalb viele Aliens an Essen denken, weil sie sich alle in einem Restaurant befinden, gerade Mittagszeit ist und daher auch das Zielalien an Essen denken sollte). Konsequenterweise treten die Abhängigkeiten für beide Zustände der direkten Ursache in ziemlich genau gleicher Höhe auf, wie man es für eine gemeinsam auf die Variablen wirkenden, generativen Ursache erwarten würde.<sup>45</sup> Dieses Problem, das an der Grundaussage des Modells aber nichts ändert, soll in der Gesamtdiskussion in Kapitel 7 nochmals aufgegriffen und in größerem Zusammenhang diskutiert werden.

---

<sup>45</sup> Ähnliche Ergebnisse, also ebenfalls Markov-verletzende Bewertungen, fanden auch Rehder und Burnett (2005) in einem Kettenexperiment.

### 5.3 Schließen im Fall weiterer beobachteter präventiver Ursachen

Im vorhergehenden Abschnitt wurde gezeigt, dass das Modell auch auf die Struktur einer kausalen Kette angewendet werden kann und dass in einem solchen Fall sowohl modelltechnisch als auch empirisch der Einfluss der Annahmen über die zugrundeliegenden Kausalprozesse verschwindet. Damit wurde ein weiterer, struktureller Beleg erbracht, dass es sich bei dem gefundenen Unterschied zwischen einer „*Reading*“- und einer „*Sending*“-Bedingung wirklich um eine an die Ursache-Effekt-Relation gekoppelte Fehlerattribution handelt.

Wenn nun der zugrundeliegende Prozess mithin wirklich mit einer Fehlerattribution erklärt werden kann, dann sollte sich diese Attribution auch in den zuvor verwendeten *Common-Cause*-Strukturen strukturell beeinflussen lassen, in dem weitere Variablen in das kausale System eingeführt werden, die für die Abwesenheit der beobachtbaren Effekte verantwortlich gemacht werden können. Stellt man sich z.B. eine *Common-Cause*-Struktur vor, die mit einem kausalen Prozess beschrieben wird, der eine ursacheseitige Fehlerattribution nahelegt (z.B. „*Sending*“), dann wird normalerweise die Abwesenheit mehrerer Effekte der anwesenden gemeinsamen Ursache auf die gemeinsame Fehlerquelle attribuiert. Die Einschätzung der Anwesenheit des Zieleffekts sollte in einer solchen Situation entsprechend niedrig ausfallen, da auch dieser betroffen ist. Führt man in dieses System nun eine weitere präventive, als anwesend beobachtete Ursache ein, die nur die beobachteten Effekte, aber nicht den unbeobachteten Effekt betrifft, dann sollte die Abwesenheit der beobachteten Effekte vornehmlich auf die als anwesend beobachtete, weitere präventive Ursache und weniger auf die nicht-beobachtete Fehlerquelle attribuiert werden. In einem solchen Fall sollte das Urteil über die Anwesenheit des Zieleffekts also deutlich weniger von der Anzahl der als abwesend beobachteten Effekte abhängig sein.

Im vorliegenden Abschnitt sollen daher die Auswirkungen der Einführung einer solchen weiteren, beobachteten Fehlerquelle auch formal auf Modellebene untersucht, Vorhersagen abgeleitet und in einem Experiment getestet werden.

#### 5.3.1 Modellierung

Ausgangspunkt für die Modellierung ist wieder eine *Common-Cause*-Struktur mit einer Ursache  $C$  und den Effekten  $E_1, \dots, E_n$ , wie sie in Kapitel 4 eingeführt worden

ist (die detaillierte Herleitung findet sich in Anhang H). Sowohl die Ursache  $C$  als auch die Effekte  $E_1, \dots, E_{n-1}$  seien beobachtet. Effekt  $E_n$  sei unbeobachtet, und Aufgabe ist es, die bedingte Wahrscheinlichkeit seiner Anwesenheit zu bestimmen. Zur Erklärung von Markov-Verletzungen wurde im Basismodell in die *Common-Cause*-Struktur eine gemeinsame Fehlerquelle  $PN$  eingeführt, die unbeobachtet ist und im Falle ihrer Anwesenheit die Kausalstärke aller Ursache-Effekt-Relation gleichermaßen senkt. Die Stärke dieser Fehlerquelle,  $w_{PN}$ , beschreibt die Annahmen der Versuchspersonen hinsichtlich der zugrundeliegenden Kausalprozesse. Fügen wir diesem Modell nun eine weitere, analog wirkende gemeinsame Fehlerquelle  $PN^*$  mit Stärke  $w_{PN^*}$  hinzu (siehe Abbildung 28), deren Zustand jedoch beobachtet ist und welche nur auf die beobachteten Effekte wirkt, lässt sich die Zielinferenz analog zu Abschnitt 4.2 (näheres siehe dort) beschreiben als:

$$P(E_n = 1 | C, PN^*, E_1, \dots, E_{n-1}; \omega) = \iint P(E_n = 1 | C, PN^*, E_1, \dots, E_{n-1}; w) \cdot P(w | \omega) dw \quad (20)$$

mit

$$P(E_n = 1 | C, PN^*, E_1, \dots, E_{n-1}; w) = \sum_{PN} P(E_n = 1 | C, PN, PN^*, E_1, \dots, E_{n-1}; w) \cdot P(PN | C, PN^*, E_1, \dots, E_{n-1}; w) \quad (21)$$

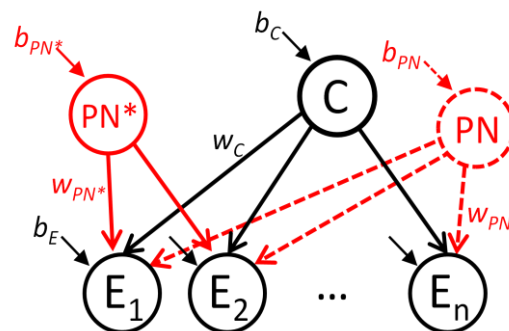


Abbildung 28. Das um eine weitere, allerdings beobachtbare gemeinsame Fehlerquelle  $PN^*$  erweiterte Basismodell. Die beobachtbare Fehlerquelle (präventive Ursache) wirkt nur auf die beobachteten Effekte  $E_1, \dots, E_{n-1}$ , nicht jedoch auf den nicht beobachteten Zieleffekt  $E_n$ .

Da der Zieleffekt  $E_n$  gegeben  $C$  und  $PN$  unabhängig von den weiteren Effekten als auch unabhängig von der zusätzlichen Fehlerquelle  $PN^*$  ist (da diese nur

die anderen Effekte beeinflusst), vereinfacht sich die auf  $PN$  bedingte Zielinferenz analog zu Gleichung 11 in Abschnitt 4.2 entsprechend:

$$P(E_n = 1|C, PN^*, E_1, \dots, E_{n-1}; w) = \sum_{PN} P(E_n = 1|C, PN; w) \cdot P(PN|C, PN^*, E_1, \dots, E_{n-1}; w) \quad (22)$$

Es lässt sich leicht erkennen, dass der zugrundeliegende Prozess wieder abstrakt mit zwei Stufen beschrieben werden kann. Nur der erste Schritt, nämlich die Schätzung der Wahrscheinlichkeit der Anwesenheit der gemeinsamen Fehlerquelle  $PN$ , also die Fehlerattribution, ist vom Zustand der beobachteten Fehlerquelle  $PN^*$  beeinflusst. Der zweite Schritt, also die Bestimmung der Wahrscheinlichkeit der Anwesenheit des Zieleffekts gegeben der An- oder Abwesenheit der Fehlerquelle  $PN$  ist von  $PN^*$  unbeeinflusst, da  $PN^*$  nur auf die Effekte  $E_1, \dots, E_{n-1}$  wirkt.

Damit sollte die Vorhersage im Falle der Abwesenheit der beobachtbaren Fehlerquelle der einer Struktur ohne diese Fehlerquelle entsprechen. In Abbildung 29 ist die Modellvorhersage dargestellt. Hierfür wurden die Parameter aus den gleichen Verteilungen gezogen, wie sie für die Simulation der Vorhersagen des Basismodells (Abschnitt 4.3) verwendet wurden; die Stärke der nicht beobachteten Fehlerquelle ( $w_{PN}$ ) wurde aus einer  $Beta(10,1)$ -Verteilung mit dem Erwartungswert 0.91 gezogen (starke gemeinsame Fehlerquelle), was im Basismodell der „*Sending*“-Bedingung entspricht. Die Vorhersage für den Fall der Abwesenheit der beobachtbaren Fehlerquelle ( $PN^* = 0$ ) entspricht damit der Vorhersage des Basismodells für die „*Sending*“-Bedingung (siehe v.a. die gestrichelte der oberen drei Linien in Abbildung 29 im Vergleich zu Abbildung 16 auf Seite 64); die Einschätzung der Wahrscheinlichkeit der Anwesenheit des Zieleffekts hängt sehr stark von der Anzahl der weiteren als anwesend beobachteten Effekte ab. Im Fall der Anwesenheit der beobachtbaren Fehlerquelle ( $PN^* = 1$ ) nimmt diese Abhängigkeit ab, und zwar umso stärker, je stärker die beobachtbare Fehlerquelle  $PN^*$  ist (vergleiche dazu in Abbildung 29 die gepunktete Linie [wenig starke beobachtete Fehlerquelle] mit der durchgezogenen Linie [starke beobachtete Fehlerquelle]).

Als Vorhersage des Modells ergibt sich damit, dass im Fall der Anwesenheit der gemeinsamen Ursache in einer *Common-Cause*-Struktur mit einer als stark wirkend gedachten gemeinsamen unbeobachteten Fehlerquelle (z.B. „*Sending*“) die Wahrscheinlichkeit der Anwesenheit eines unbeobachteten Zieleffekts stark von der Anzahl der weiteren als anwesend beobachteten Effekte abhängig ist, sofern keine weitere präventive Ursache anwesend ist (Vorhersage entspricht Basismodell). Ist jedoch eine weitere präventive Ursache als anwesend beobachtet und wirkt diese nicht auf den Zieleffekt, aber auf alle anderen Effekte der *Common-Cause*-Struktur, dann sollte die Abhängigkeit der Wahrscheinlichkeit der Anwesenheit des Zieleffekts deutlich geringer ausfallen. Im Fall der Abwesenheit der Ursache sollte es keine Unterschiede geben.

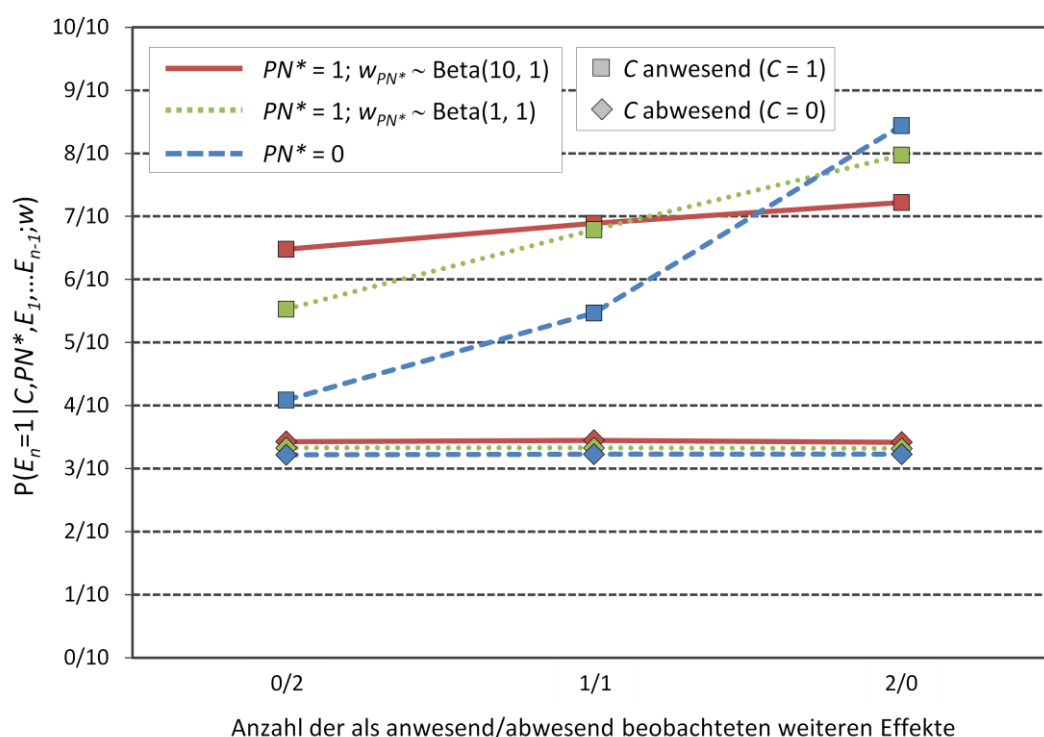


Abbildung 29. Modellvorhersage für die Abwesenheit ( $PN^* = 0$ ) und Anwesenheit ( $PN^* = 1$ ) der beobachteten präventiven Ursache  $PN^*$  und für deren unterschiedliche Stärke ( $w_{PN^*}$ ) auf der Basis einer Monte-Carlo-Simulation mit 100 000 Durchgängen in einem *Common-Cause*-Modell mit einer Ursache  $C$  und drei Effekten für den Fall einer starken unbeobachteten gemeinsamen Fehlerquelle  $PN$  („*Sending*“-Fall;  $w_{PN} \sim \text{Beta}(10,1)$ ). Die weiteren Parameter des kausalen Systems wurden hierfür aus folgenden Verteilungen gezogen:  $w_C \sim \text{Beta}(10,1)$ ,  $b_E \sim \text{Beta}(1,2)$  und  $b_{PN} \sim \text{Beta}(1,1)$ .

Die Vorhersagen für  $C = 0$  liegen übereinander; die drei Linien wurden daher mit einem Versatz (+0.1, 0, -0.1) versehen, um sie optisch sichtbar zu machen.

### 5.3.2 Experiment 6: Überblick

Die Vorhersage, dass die konditionale Abhängigkeit der Einschätzung der Anwesenheit eines unbeobachteten Effekts in einer *Common-Cause*-Struktur bei Anwesenheit der gemeinsamen Ursache abnimmt, wenn eine beobachtete Fehlerquelle hinzugefügt wird, die (nur) auf die beobachteten Effekte wirkt, soll im folgenden Experiment empirisch getestet werden. Hierfür bietet sich wieder das Paradigma der Gedankenlesenden-vs.-Gedankensendenden Aliens an, insbesondere die „*Sending*“-Bedingung. In dieser konnte empirisch eine starke Abhängigkeit der Einschätzung der Anwesenheit des Zieleffekts von der Anzahl der weiteren als anwesend beobachteten Effekte gezeigt werden. Führt man in ein solches System eine beobachtbare Fehlerquelle ein, die präventiv nur auf die beobachteten Effekte wirkt, sollte eine eventuelle Abwesenheit dieser Effekte bei Anwesenheit der gemeinsamen Ursache auf die neue Fehlerquelle attribuiert werden. Da diese nicht auf den Zieleffekt wirkt, sollte dessen Einschätzung in einem solchen Fall weniger stark vom Zustand der anderen Effekte abhängig sein.

Als mögliche beobachtbare, präventive Ursache im „*Sending*“-Fall kann eine Wand fungieren, die zwischen dem sendenden Alien und den beobachteten Effekten, aber nicht dem unbeobachteten Effekt platziert ist. Da dicke Wände dafür bekannt sind, Funkübertragungen u.Ä. zu stören, sollte – entsprechende Intuitionen der Probanden vorausgesetzt – in einem solchen Fall die Einschätzung, ob das Zielalien an „POR“ denkt, weniger davon abhängen, wie viele der anderen Effektalien an „POR“ oder an nichts denken. Die entsprechende Vergleichsbedingung wäre eine Bedingung ohne Wände (entspricht einer Replikation der „*Sending*“-Bedingung in den Basisexperimenten). Als weitere Vergleichsbedingung bietet sich eine Bedingung mit Wänden an, die mit offenen Türen versehen sind und damit keinen präventiven Charakter im Hinblick auf die Gedankenübertragung haben. Letztere Manipulation wäre insoweit subtiler, als der Unterschied zwischen der „Offenen-Türen“-Bedingung und der „Wände“-Bedingung lediglich der präventive Charakter des eingeführten Objekts ist (und nicht die Einführung eines Objektes an sich). Der gefundene Unterschied zwischen den Bedingungen könnte damit sauberer auf eben diese Eigenschaft als präventive Ursache zurückgeführt werden.

### 5.3.3 Experiment 6: Methode

*Versuchspersonen.* 48 Personen nahmen an der Untersuchung teil (26 Frauen, 22 Männer; Durchschnittsalter 24.1 Jahre, Altersspanne 19–56 Jahre). Die Anwerbung erfolgte im Zentralen Hörsaalgebäude (ZHG), auf dem Campus, im Psychologischen Institut, in den Mensen rund um den Campus der Universität Göttingen und zum Teil im Bekannten- und Verwandtenkreis der Versuchsleiterinnen und Versuchsleiter. Es handelte sich dabei größtenteils um Studierende der Universität Göttingen verschiedenster Fachrichtungen. Keiner der Probanden hatte bereits an einem der vorhergehenden Experimente der Serie teilgenommen. Die Versuchspersonen nahmen freiwillig an der Untersuchung teil. Nach Ende des Experiments wurden die Teilnehmerinnen und Teilnehmer mit einer Süßigkeit belohnt und sofern gewünscht über den Zweck und das Ziel der Untersuchung genauer aufgeklärt. Die Probanden wurden den drei Bedingungen („Wände“, „Keine Wand“ und „Offene Türen“) zufällig zugeteilt ( $N = 3 \times 16$ ).

*Versuchsdurchführung und Material.* Die Durchführung von Experiment 6 entsprach im Wesentlichen der Durchführung der vorangegangenen Experimente. Das Experiment bestand aus einer Instruktions- und einer Testphase. In der *Instruktionsphase* wurde den Probanden eine vierseitige Instruktion vorgelegt. Den Probanden wurde mitgeteilt, dass es vier Aliens – Gonz, Zoohng, Murks und Bruuf – gebe, die meistens an nichts und manchmal an „POR“ denken würden (Bildmaterial siehe Abbildung 30). Weiter wurde erklärt, dass Gonz seine Gedanken aussenden und damit in die Köpfe der anderen Aliens übertragen könne. Des Weiteren wurden wieder am Beispiel der Relation Gonz und Murks die vier möglichen Fälle erklärt. Der Text entsprach dem Text für die „Sending“-Bedingung aus Experiment 5. Die Instruktion war für alle Probanden identisch.



Abbildung 30. Die Abbildungen der vier Außerirdischen Gonz, Zoohng, Murks und Bruuf, wie sie in der Instruktion von Experiment 6 eingeführt wurden. Das Material wurde größtenteils von Steyvers et al. (2003) übernommen.

Am Ende der Instruktionsphase, nach dem Lesen der Versuchsinstruktion, hatten die Probanden die Instruktion in ihren wesentlichen Inhalten noch einmal wiederzugeben, um diese dann ggf. nochmals zu lesen.

In der *Testphase* wurden den Probanden wieder je sechs Testfälle in randomisierter Reihenfolge vorgelegt, in denen die Wahrscheinlichkeit der Anwesenheit des Zieleffekts, hier immer Murks, auf einer Skala von 0 bis 10 einzuschätzen war – gegeben der Gedanken von Gonz (Ursache) und der Gedanken von Zoohng und Bruuf (Effekte). Die Testfälle waren wieder derart gestaltet, dass jede Versuchsperson je Zustand der Ursache (nichts vs. „POR“) die drei möglichen Fälle für die Anzahl der als anwesend beobachteten weiteren Effekte (0, 1 oder 2) vorgelegt bekam. Die drei experimentellen Bedingungen unterschieden sich nur in der Anwesenheit von präventiven Ursachen. Die „Keine-Wände“-Bedingung entsprach dabei der ursprünglichen „*Sending*“-Bedingung in den Basisexperimenten, außer dass die Aliens sich in einem großen Raum befanden (siehe für einen Beispiel-Testfall Abbildung 31).

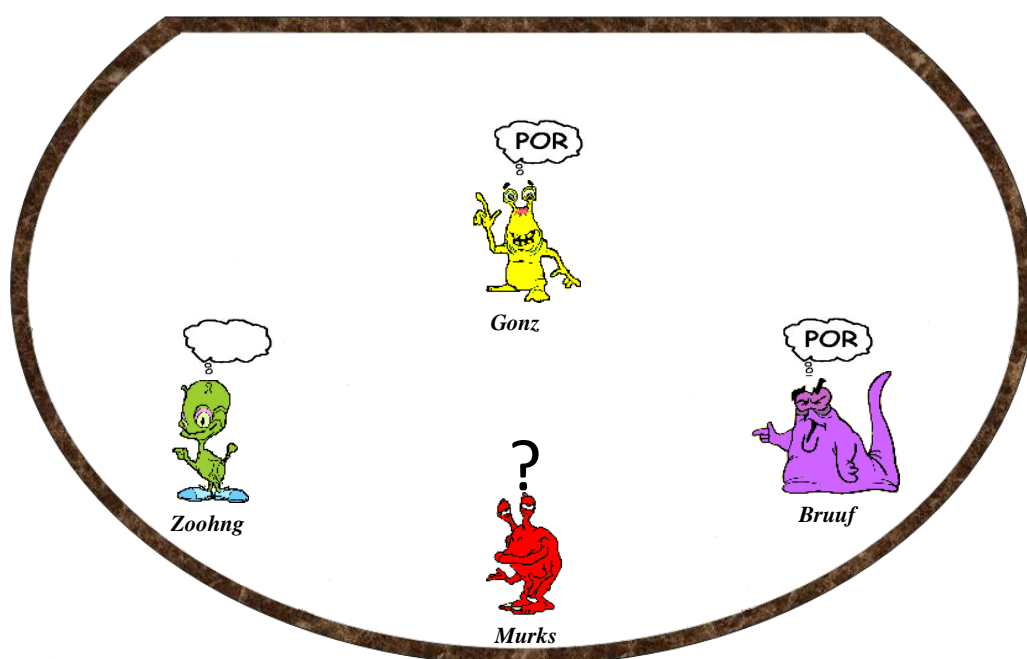


Abbildung 31. Ein Beispiel-Testfall, wie er in der Testphase von Experiment 6 in der „Keine-Wände“-Bedingung benutzt wurde. Die Versuchspersonen hatten sich zehn Situationen mit der gegebenen Konfiguration vorzustellen und dann anzugeben, in wie vielen dieser Situationen das Zielalien (immer Murks) an „POR“ denken würde (Skala von 0 bis 10).



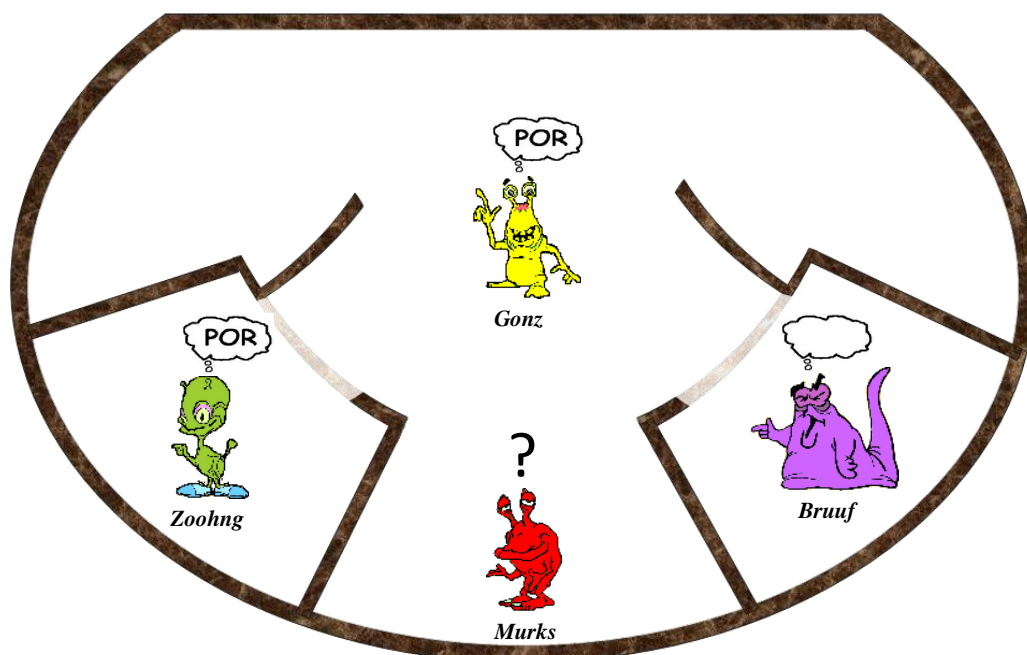


Abbildung 32. Ein Beispiel-Testfall, wie er in der Testphase von Experiment 6 in der „Offene-Türen“-Bedingung benutzt wurde. Die Probanden hatten anzugeben, in wie vielen von zehn Situationen mit der Gedankenkonfiguration das Zielalien (immer Murks) an „POR“ denken würde (Skala von 0 bis 10).

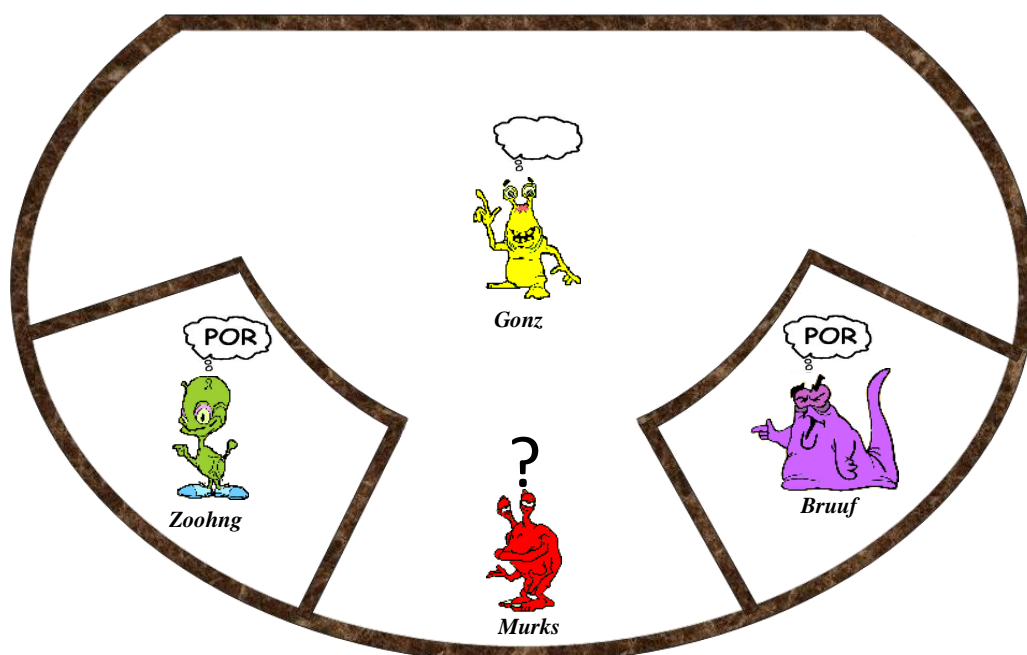


Abbildung 33. Ein Beispiel-Testfall, wie er in der Testphase von Experiment 6 in der „Wände“-Bedingung benutzt wurde. Die Probanden hatten anzugeben, in wie vielen von zehn Situationen mit der Gedankenkonfiguration das Zielalien (immer Murks) an „POR“ denken würde (Skala von 0 bis 10).

In der „Offene-Türen“-Bedingung (für einen Beispiel-Testfall siehe Abbildung 32) waren Wände in den Raum eingefügt, die jedoch zwischen den Effekt-

aliens und dem Gedankensender durchbrochen waren. Den Versuchspersonen wurde dazu im Text mitgeteilt, dass Gonz, Zoohng, Bruuf und Murks alle in einem großen Raum seien. Zoohng und Bruuf saßen je in einem Zimmer. Die Türen zwischen ihnen und Gonz seien aber offen, so dass sie Gonz sehen könne. Durch diesen Hinweis sollte sichergestellt werden, dass den Probanden klar ist, dass die eingefügten Wände aller Wahrscheinlichkeit nach die Gedankenübertragung nicht beeinträchtigen würden.

In der „Wände“-Bedingung (für einen Beispiel-Testfall siehe Abbildung 33) waren Wände in den Raum eingefügt, ohne dass es eine Öffnung gab. Die Bedeutung der Wände und ihre Wirkung auf die Gedankenübertragung wurden offen gelassen. Im Aufgabentext wurde lediglich explizit auf die Wände hingewiesen: Gonz, Zoohng, Bruuf und Murks seien alle in einem großen Raum. Zoohng und Bruuf saßen aber je in einem Zimmer, so dass eine Wand zwischen ihnen und Gonz sei.

Für die Durchführung des Versuchs wurde ein Aktenordner angelegt, der die Instruktion (vier Seiten) und die 24 möglichen Testfälle<sup>46</sup> sowie die Daten- und eine mit Excel erzeugte Randomisierungstabelle enthielt. Die Durchführung erfolgte auf dem Campus der Universität Göttingen sowie in umliegenden Universitätsgebäuden und dauerte ca. 10–15 Minuten pro Versuchsperson.

*Design.* Im Experiment sollen sowohl die beiden experimentellen Kontrollbedingungen („Offene Türen“ vs. „Keine Wände“) verglichen werden und, sofern kein Unterschied gefunden wird, ebenso die „Wände“-Bedingung mit der „Offene Türen“-Bedingung. Daraus ergeben sich zwei 2x2x3-Designs mit je einem experimentellen Faktor (Faktor *EXT*: „Offene Türen“ vs. „Keine Wände“ bzw. „Offene Türen“ vs. „Wände“) und je zwei Messwiederholungsfaktoren, mithin der Status der Ursache *C* (anwesend vs. abwesend) und der Anzahl der als anwesend beobachteten weiteren Effekte (0, 1 oder 2; Faktor  $N_{E+}$ ).

---

<sup>46</sup> Je Zustand der Ursache (nichts vs. „POR“) ein Testfall für „kein Effektalien denkt an Essen“, zwei Testfälle für „ein Effektalien denkt an Essen“ (entweder Zoohng oder Bruuf; welches Item vorgelegt wurde, wurde zufällig bestimmt) und ein Testfall für „beide Effektaliens denken an Essen“; also 4 je Zustand der Ursache, für jede der drei experimentellen Bedingungen demnach 8, insgesamt also 24.

*Vorhersage.* Für den Vergleich der beiden Kontrollbedingungen („Offene Türen“ vs. „Keine Wände“) ergibt sich als Vorhersage, dass es keinen Unterschied bezüglich des experimentellen Faktors (Faktor *EXT*) geben sollte, sowohl nicht im Hinblick auf Haupteffekte (Haupteffekt Faktor *EXT*) als auch im Hinblick auf Interaktionseffekte (Interaktionseffekte  $EXT \times C$ ,  $EXT \times N_{E+}$  als auch  $EXT \times C \times N_{E+}$ ), da es in beiden Fällen keine beobachtbar anwesende präventive Ursache gibt.

Für den Vergleich der „Offene-Türen“- mit der „Wände“-Bedingung ergibt sich eine Vorhersage, die im Wesentlichen der Vorhersage der Basisexperimente entspricht. Im Falle der Abwesenheit der Ursache sollte sich sowohl kein Einfluss des Status der anderen Effekte (Faktor  $N_{E+}$ ) als auch keine Interaktion dieses Faktors mit der experimentellen Manipulation zeigen. Im Falle der Anwesenheit der Ursache sollte sich in der „Offene-Türen“-Bedingung jedoch eine starke Abhängigkeit der Bewertungen von der Anzahl der als anwesend beobachteten weiteren Effekte ergeben, die in der „Wände“-Bedingung deutlich schwächer ausgeprägt sein sollte (Interaktionseffekt  $EXT \times N_{E+}$  für  $C=1$ ), da in dieser dem Modell nach die Abwesenheit der weiteren Effekte auf die Anwesenheit der eingeführten präventiven Ursache (die Wände) attribuiert wird. Es wird ebenfalls eine Drei-Wege-Interaktion erwartet.

#### 5.3.4 Experiment 6: Ergebnisse

Die Ergebnisse von Experiment 4 sind in Abbildung 34 sowie in Tabelle 8 dargestellt; die detaillierten varianzanalytischen Ergebnisse finden sich in Tabelle I-1, Tabelle I-2, Tabelle I-3 und Tabelle I-4 in Anhang I.

*Auswertung im Hinblick auf die Kontrollbedingungen („Offene Türen“ vs. „Keine Wände“).* Die Bewertungen der Wahrscheinlichkeit der Anwesenheit des unbeobachteten Zieleffekts sind im Durchschnitt deutlich höher, wenn die Ursache ebenfalls anwesend ist (Haupteffekt  $C$ :  $F_{1,30}=71.91$ ,  $p<.001$ ,  $\eta^2=.71$ ). Ein Einfluss der experimentellen Manipulation, der für die beiden Kontrollbedingungen auch nicht erwartet worden ist, zeigt sich nicht. Die Bewertungen liegen insbesondere im Falle der Anwesenheit der Ursache praktisch direkt aufeinander (siehe die unteren beiden der oberen drei Linien in Abbildung 34). Es gibt sowohl keinen Haupteffekt der Manipulation selbst (Haupteffekt  $EXT_{Control}$ :  $F_{1,30}<1$ ,  $p=.96$ ) als

auch keine Zwei-Wege-Interaktionen (Interaktionseffekt  $EXT_{Control} \times C$ :  $F_{1,30} < 1$ ,  $p = .86$ ; Interaktionseffekt  $EXT_{Control} \times N_{E+}$ :  $F_{2,60} < 1$ ,  $p = .84$ ). Eine Drei-Wege-Interaktion konnte ebenfalls nicht entdeckt werden (Interaktionseffekt  $EXT_{Control} \times C \times N_{E+}$ :  $F_{2,60} < 1$ ,  $p = .78$ ).

*Auswertung im Hinblick auf die eigentliche experimentelle Manipulation („Offene Türen“ vs. „Wände“).* Beim Vergleich der beiden relevanten experimentellen Bedingungen findet sich wieder, dass die Einschätzung der Anwesenheit des Zieleffekts im Durchschnitt deutlich höher ist, wenn die Ursache anwesend als wenn diese abwesend ist (Haupteffekt  $C$ :  $F_{1,30} = 77.09$ ,  $p < .001$ ,  $\eta^2 = .72$ ).

Im Fall der Anwesenheit der Ursache (Gonz denkt an „POR“; die durchgezogene und die gestrichelte der drei oberen Linien in Abbildung 34) hängt die Wahrscheinlichkeitseinschätzung bezüglich der Anwesenheit des Zieleffekts deutlich von der Anzahl der weiteren als anwesend („POR“) beobachteten Effekte ab (Haupteffekt  $N_{E+}$  für  $C=1$ :  $F_{2,60} = 13.82$ ,  $p < .001$ ,  $\eta^2 = .32$ ). Ausgehend von null als anwesend beobachteten Effekten steigt die Einschätzung mit deren zunehmender Anzahl in der „Offene-Türen“-Bedingung von 4.94 auf 7.81 ( $F_{2,30} = 8.27$ ,  $p < .01$ ,  $\eta^2 = .36$ ) und in der „Wände“-Bedingung von 6.13 auf 8.13 ( $F_{2,30} = 5.58$ ,  $p < .01$ ,  $\eta^2 = .27$ ). Der Anstieg in der „Wände“-Bedingung ist damit zwar tendenziell wie vorhergesagt etwas schwächer als in der „Offene-Türen“-Bedingung, eine Interaktion kann aber nicht nachgewiesen werden (Interaktionseffekt  $EXT \times N_{E+}$  für  $C=1$ :  $F_{2,60} < 1$ ,  $p = .63$ ).

Ist die Ursache abwesend (Gonz denkt an nichts; die unteren Linien in Abbildung 34), dann schwanken die Bewertungen relativ unsystematisch zwischen 2.75 und 3.56. Weder hat die Anzahl der als anwesend beobachteten Effekte einen signifikanten Einfluss auf die Beurteilungen (Haupteffekt  $N_{E+}$  für  $C=0$ :  $F_{2,60} < 1$ ,  $p = .91$ )<sup>47</sup>, noch deren Interaktion mit dem „Offene-Türen“-vs.-„Wände“-Faktor (Interaktionseffekt  $EXT \times N_{E+}$  für  $C=0$ :  $F_{2,60} = 1.18$ ,  $p = .31$ ).

Eine Drei-Wege-Interaktion kann ebenfalls nicht nachgewiesen werden (Interaktionseffekt  $EXT \times C \times N_{E+}$ :  $F_{2,60} = 1.32$ ,  $p = .27$ ).

<sup>47</sup> Dies gilt sowohl für die „Offene-Türen“-Bedingung ( $F_{2,30} < 1$ ,  $p = .67$ ) als auch für die „Wände“-Bedingung ( $F_{2,30} = 1.72$ ,  $p = .20$ ).

Tabelle 8

## Durchschnittliche Bewertung der Anwesenheit des Zieleffekts (Experiment 6)

	Kontrolle – Keine Wände $EXT = 0$ ( $n = 16$ )			Kontrolle – offene Türen $EXT = 0$ ( $n = 16$ )			Wände $EXT = 1$ ( $n = 16$ )		
	$N_{E+} = 0$	$N_{E+} = 1$	$N_{E+} = 2$	$N_{E+} = 0$	$N_{E+} = 1$	$N_{E+} = 2$	$N_{E+} = 0$	$N_{E+} = 1$	$N_{E+} = 2$
$C = 1$ („POR“)	5.06 (3.15)	6.31 (2.02)	8.00 (1.93)	4.94 (2.69)	6.38 (2.25)	7.81 (1.94)	6.13 (2.78)	7.31 (1.66)	8.13 (1.20)
$C = 0$ (nichts)	3.13 (1.26)	3.31 (1.40)	3.44 (1.36)	3.56 (2.22)	3.25 (1.98)	3.19 (1.60)	2.75 (1.98)	2.88 (2.00)	3.13 (2.36)

Anmerkungen. Die Tabelle zeigt die durchschnittlichen Bewertungen auf einer Skala von „0 von 10“ bis „10 von 10“ (Häufigkeitsskala mit 11 Stufen) bezüglich der Wahrscheinlichkeit der Anwesenheit des Zieleffekts (Standardabweichung in Klammern).  $C$  = Zustand der gemeinsamen Ursache;  $N_{E+}$  = Anzahl der als anwesend beobachteten weiteren Effekte;  $EXT$  = Zustand der beobachteten präventiven Ursache (abwesend [0] vs. anwesend [1]).

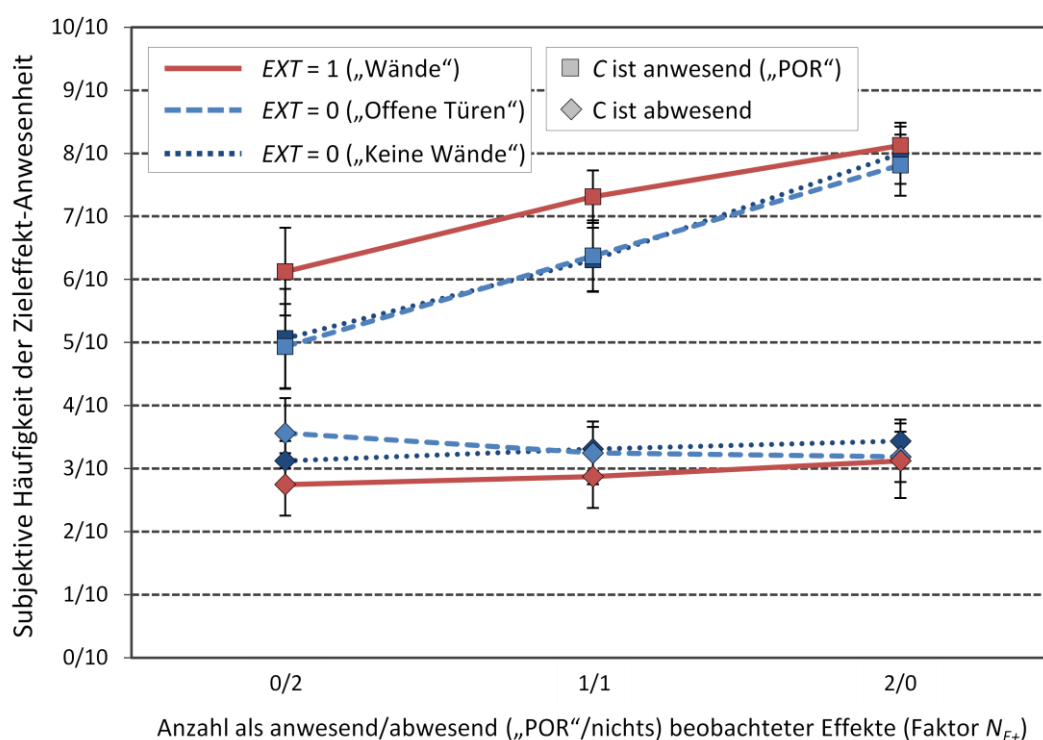


Abbildung 34. Durchschnittliche Ratings bezüglich der Häufigkeit der Anwesenheit des Zieleffekts in zehn fiktiven Situationen gegeben des Status der gemeinsamen Ursache  $C$  und der Anzahl der als anwesend beobachteten weiteren Effekte (Faktor  $N_{E+}$ ; jeweils intraindividuelle Variation) über den Faktor der externen präventive Ursache (Faktor  $EXT$ : interindividuelle Manipulation,  $N = 3 \times 16$ ) in Experiment 6. Die Fehlerbalken stellen den Standardfehler des Mittelwerts dar.

### 5.3.5 Experiment 6: Diskussion

Auch wenn eine der Tendenz nach geringere Abhängigkeit der Beurteilungen der Wahrscheinlichkeit der Anwesenheit des Zieleffekts von der Anzahl der als anwesend beobachteten weiteren Effekte in der „Wände“-Bedingung im Vergleich zur „Offene-Türen“-Bedingung zu beobachten war, konnte dies jedoch statistisch nicht abgesichert werden. Ein Grund hierfür mag sein, dass den Probanden nicht mitgeteilt wurde, ob die Wände einen Einfluss auf die Gedankenübertragung haben und dass dies durch die interindividuelle Manipulation auch nicht offensichtlich war. Den Probanden in einer Bedingung wurden nur Testfälle vorgelegt, die entweder Wände, offene Türen oder keine Wände enthielten. Da über diese keine Angaben gemacht wurden und der Vergleich fehlte, mögen sich die Versuchspersonen über deren Bedeutung keine Gedanken gemacht haben. Bevor eine Aussage über das Modell getroffen werden kann, soll es daher in abgeänderter Form nochmals getestet werden.

### 5.3.6 Experiment 7: Überblick

In Experiment 7 soll die Vorhersage des Modells nochmals getestet werden. Um die mögliche Wirkung der Wände als präventive Ursache herauszustellen, soll diesmal der Faktor *EXT* intraindividuell manipuliert werden. Das heißt, jede Versuchsperson sieht Testfälle die verschiedene Realisierungen des Faktors *EXT* enthalten. Um dabei die Belastung der Versuchspersonen in der Testphase zu reduzieren, sollen dazu nur die beiden Bedingungen „Wände“ und „Offene Türen“ zum Einsatz kommen und auch auf die Fälle, in denen die Ursache abwesend ist, verzichtet werden. Dies erscheint vertretbar, da die relevante Modellvorhersage sich auf die Interaktion bei Anwesenheit der Ursache bezieht.

Die Manipulation des Faktors *EXT* innerhalb der Versuchspersonen soll diese dazu anregen, über die Bedeutung des Unterschiedes zwischen der „Wände“- und der „Offenen-Türen“-Bedingung nachzudenken, und damit eine mögliche präventive Wirkung der Wände herausstellen.

### 5.3.7 Experiment 7: Methode

*Versuchspersonen.* 24 Personen nahmen an der Untersuchung teil (14 Frauen, 10 Männer; Durchschnittsalter 26.0 Jahre, Altersspanne 20–43 Jahre). Diese wurden

auf dem Campus, im Zentralen Hörsaalgebäude (ZHG), im Psychologischen Institut, in den Mensen rund um den Campus der Universität Göttingen und zum Teil im Bekannten- und Verwandtenkreis der Versuchsleiterinnen und Versuchsleiter angeworben. Es handelte sich dabei größtenteils um Studierende der Universität Göttingen verschiedenster Fachrichtungen. Keiner der Probanden hatte bereits an einem der vorhergehenden Experimente der Serie teilgenommen. Die Teilnahme erfolgte freiwillig. Nach Ende der Untersuchung wurden die Teilnehmerinnen und Teilnehmer mit einer Süßigkeit belohnt und sofern gewünscht über den Zweck und das Ziel der Untersuchung genauer aufgeklärt.

*Versuchsdurchführung und Material.* Das Experiment bestand wieder aus einer Instruktions- und einer Testphase. Die *Instruktionsphase* war identisch mit der von Experiment 6; es wurde auch dieselbe Instruktion verwendet.

In der Testphase wurden den Probanden sechs Testfälle in zufälliger Reihenfolge vorgelegt. Das Material für die Testfälle wurde von Experiment 6 übernommen. Die Testfälle wurden derart ausgewählt, dass je Versuchsperson für jede der beiden Bedingungen des Faktors *EXT* („Offene Türen“ und „Wände“) für den Faktor  $N_{E+}$  die drei relevanten Ausprägungen abgedeckt waren: Keines der anderen Effektaliens denkt an „POR“, eins von den anderen Effektaliens denkt an „POR“ (welches wurde für jede Versuchsperson zufällig bestimmt) und beide Effektaliens denken an „POR“. Das Alien, das die gemeinsame Ursache darstellte (Gonz), dachte immer an „POR“; die Ursache war mithin immer anwesend.

Für die Durchführung des Versuchs wurde ein Aktenordner angelegt, der die Instruktion (vier Seiten), die acht möglichen Testfälle sowie die Daten- und eine mit Excel erzeugte Randomisierungstabelle enthielt. Die Durchführung erfolgte auf dem Campus der Universität Göttingen sowie in umliegenden Universitätsgebäuden und dauerte ca. 10–15 Minuten pro Versuchsperson.

*Design.* Da nur die Fälle mit anwesender Ursache getestet wurden, handelt es sich um ein 2x3-Design mit zwei Messwiederholungsfaktoren: dem Faktor *EXT* (zweistufig: „Offene Türen“ vs. „Wände“) und dem Faktor  $N_{E+}$ , also der Anzahl der als anwesend beobachteten weiteren Effekte (dreistufig: 0, 1 oder 2).

*Vorhersage.* Die empirische Vorhersage leitet sich aus der Vorhersage des Modells für den Fall der Anwesenheit der Ursache ab. Hierfür wird zum einen erwartet, dass die Anzahl der als anwesend beobachteten weiteren Effekte einen deutlichen Einfluss auf die Bewertungen haben sollte (Haupteffekt  $N_{E+}$ ), und zum anderen wird erwartet, dass dieser Einfluss in der „Wände“-Bedingung deutlich schwächer ausgeprägt ist, als in der „Offene-Türen“-Bedingung (Interaktionseffekt  $EXT \times N_{E+}$ ).

### 5.3.8 Experiment 7: Ergebnisse

Die Ergebnisse von Experiment 7 sind in Tabelle 9 und in Abbildung 35 dargestellt; die detaillierten varianzanalytischen Ergebnisse finden sich in Tabelle J-1 in Anhang J.

Wie in Abbildung 35 gut zu erkennen, hängt die Wahrscheinlichkeitseinschätzung bezüglich der Anwesenheit des Zieleffekts deutlich von der Anzahl der als anwesend beobachteten weiteren Effekte ab (Haupteffekt  $N_{E+}$ :  $F_{2,46}=7.08$ ,  $p<.01$ ,  $\eta^2=.24$ ). Ausgehend von null als anwesend beobachteten weiteren Effekten steigen die Beurteilungen mit deren zunehmender Anzahl in der „Offene-Türen“-Bedingung von 4.58 auf 7.04 ( $F_{2,46}=8.39$ ,  $p<.01$ ,  $\eta^2=.27$ ) und in der „Wände“-Bedingung von 5.96 auf 7.33 ( $F_{2,46}=4.15$ ,  $p<.05$ ,  $\eta^2=.15$ ). Der Anstieg der Bewertungen ist in der „Offene-Türen“-Bedingung deutlich stärker ausgeprägt als in der „Wände“-Bedingung (Interaktionseffekt  $EXT \times N_{E+}$ :  $F_{2,46}=3.42$ ,  $p<.05$ ,  $\eta^2=.13$ ).

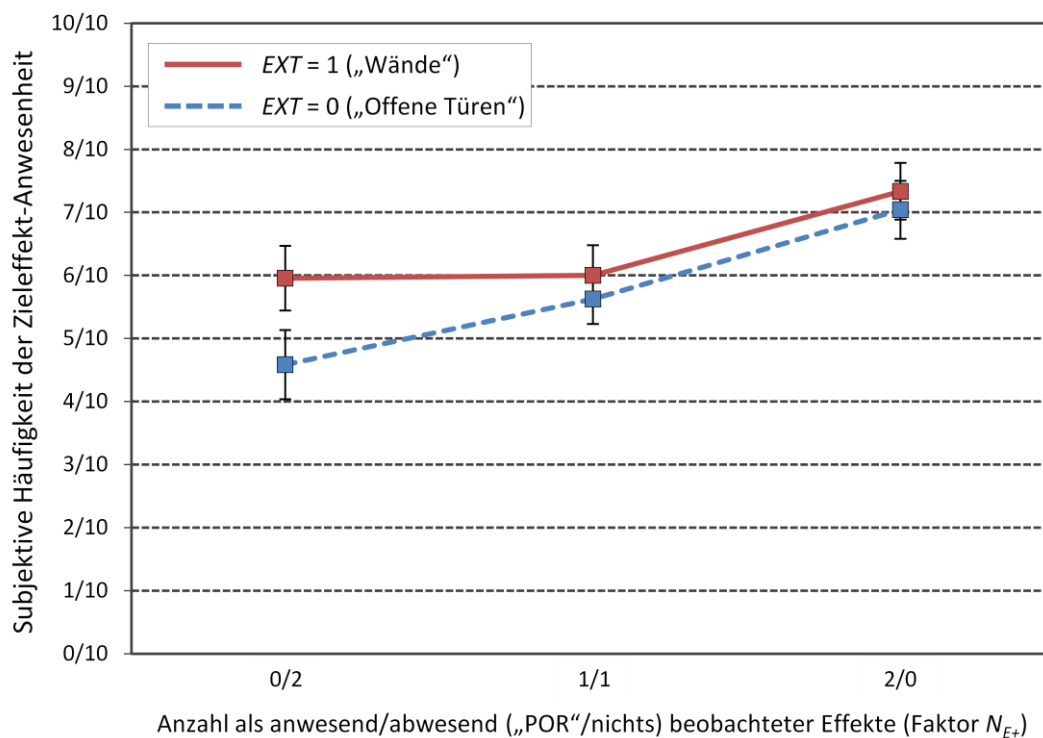
Tabelle 9

*Durchschnittliche Bewertung der Anwesenheit des Zieleffekts (Experiment 7)*

	EXT = 0 – Offene Türen			EXT = 1 – Wände		
	$N_{E+} = 0$	$N_{E+} = 1$	$N_{E+} = 2$	$N_{E+} = 0$	$N_{E+} = 1$	$N_{E+} = 2$
[C = 1] („POR“)	4.58 (2.69)	5.63 (1.93)	7.04 (2.26)	5.96 (2.51)	6.00 (2.34)	7.33 (2.20)

*Anmerkungen.* Die Tabelle zeigt die durchschnittlichen Bewertungen auf einer Skala von „0 von 10“ bis „10 von 10“ (Häufigkeitsskala mit 11 Stufen) bezüglich der Wahrscheinlichkeit der Anwesenheit des Zieleffekts (Standardabweichung in Klammern;  $N = 24$ ). C = Zustand der gemeinsamen Ursache;  $N_{E+}$  = Anzahl der als anwesend beobachteten weiteren Effekte; EXT = Zustand der beobachteten präventiven Ursache (abwesend [0] vs. anwesend [1]).





*Abbildung 35.* Durchschnittliche Ratings bezüglich der Häufigkeit der Anwesenheit des Zieleffekts in zehn fiktiven Situationen gegeben der Anwesenheit gemeinsamen Ursache  $C$  ( $C = 1$ ) und der Anzahl der als anwesend beobachteten weiteren Effekte (Faktor  $N_{E+}$ ) über den Faktor der externen präventive Ursache (Faktor  $EXT$ ; beide intraindividuelle Manipulation,  $N = 24$ ) in Experiment 7. Die Fehlerbalken stellen den Standardfehler des Mittelwerts dar.

### 5.3.9 Diskussion

In Experiment 7 konnte die Vorhersage des an den Fall weiterer beobachtbarer präventiver Ursachen angepassten Modells für den Fall der Anwesenheit der Ursache bestätigt werden. Damit scheint das Problem in Experiment 6, in dem ein statistisch bedeutsamer Unterschied zwischen der „Offene-Türen“- und der „Wände“-Bedingung im Hinblick auf deren Interaktion mit der Anzahl als anwesend beobachteter weiterer Effekte nicht gefunden werden konnte, vor allem auf die unzureichende Manipulation der präventiven Ursache zurückzuführen zu sein, nicht jedoch auf die Unangemessenheit des Modells. Insgesamt erscheint daher im Nachhinein, dass eine deutlichere, explizite Instruktion der Bedeutung der eingeführten Wände als präventive Ursache sinnvoller gewesen wäre. Die Manipulation in Experiment 6 und auch in Experiment 7 war subtiler und damit aus rein psychologischer Sicht sicherlich interessanter. Doch im Hinblick auf die Testung der Modellvorhersage war gerade nicht interessant, ob die Versuchsper-

sonen eine Wand als eine präventive Ursache für das Gedankenübertragen bei Außerirdischen ansehen – darüber macht das Modell schließlich keinerlei Vorhersagen. Getestet werden sollte, wie sich die Inferenzen der Probanden bezüglich der Anwesenheit eines unbeobachteten Zieleffekts durch die Einführung einer als anwesend beobachteten präventiven Ursache verändern. Dies setzt natürlich aber gerade voraus, dass die Versuchspersonen die Manipulation („Wände“ vs. „Offene Türen“) als eine Manipulation im Sinne des Modells, also der Anwesenheit oder Abwesenheit einer weiteren präventiv wirkenden Ursache, verstehen – und dies hätte deutlicher instruiert werden können.

Nichtsdestotrotz konnte das Modell auch für den Fall der Einführung einer beobachteten präventiven Ursache ausreichend bestätigt werden. Dies ist insbesondere deshalb interessant, da die eingeführte zusätzliche Variable direkt mit dem im Modell angenommenen Prozess der adaptiven Fehlerattribution interagiert. Der Fall liefert damit einen weiteren Beleg für die Richtigkeit der dem Modell zugrunde liegenden strukturellen Annahmen.

#### 5.4 Zusammenfassung

Im vorliegenden Kapitel wurde für das Basismodell (siehe Kapitel 4) bzw. die dem Basismodell zugrunde liegende Idee der Einführung einer gemeinsamen ursacheseitigen Fehlerquelle explizit getestet, indem das Modell in drei verschiedenen Kontexten angewendet, entsprechende Vorhersagen abgeleitet und empirisch untersucht wurden.

In Abschnitt 5.1 wurde das Modell auf eine *Common-Cause*-Struktur angewendet, in der beide Zustände der gemeinsamen Ursache kausal aktiv sind („A/B“-Fall anstatt wie im Basismodell der „0/1“-Fall, in dem nur die Anwesenheit der Ursache kausal aktiv ist). Hierfür musste lediglich die der Zielinferenz zugrunde liegende bedingte Wahrscheinlichkeitstabelle angepasst werden. In dieser Situation sagt das Modell für die Bewertung der Anwesenheit eines unbeobachteten Zieleffekts eine Abhängigkeit vom Status der weiteren Effekte für *beide* Zustände der Ursache voraus und, dass diese Abhängigkeit wie im Basismodell in einem kausalen System mit starker gemeinsamer Fehlerursache („*Sending*“-Bedingung) deutlich größer ausfallen sollte als in einem System mit schwacher gemeinsamen Fehlerursache („*Reading*“-Bedingung). Diese Vorhersa-

gen wurden experimentell in einem für den „A/B“-Fall angepassten, auf dem Paradigma der Basisexperimente beruhenden Experiment getestet. Hierfür wurden beide Gedanken der Außerirdischen („TUS“ und „POR“ statt nichts und „POR“) als kausal aktiv beschrieben und mit der bekannten „Reading“-vs.-„Sending“-Manipulation kombiniert. Wie vorhergesagt, zeigte sich eine Abhängigkeit vom Status der weiteren Effekte für beide Zustände der Ursache, und zwar deutlich stärker in der „Sending“-Bedingung.

In Abschnitt 5.2 wurde die dem Modell zugrunde liegende Idee der Einführung einer gemeinsamen, an die Ursache gebundenen Fehlerquelle auf eine *Causal-Chain*-Struktur angewendet. Für diesen Fall sagt das Modell vorher, dass der Unterschied zwischen einer „Reading“- und einer „Sending“-Bedingung verschwinden sollte, da es bei der Bewertung der potentiellen Anwesenheit des ultimativen Effekts in einer kausalen Kette nicht darauf ankommt, ob mögliche Fehler ursache- oder effektseitig attribuiert werden. Die Vorhersagen wurden experimentell dergestalt getestet, dass das in den Basisexperimenten verwendete Paradigma für eine kausale Kette angepasst wurde, indem den Probanden (entsprechend je nach Bedingung) erklärt wurde, dass in einer Reihe von Außerirdischen jeweils links stehende Alien verursache die Gedanken des jeweils rechts stehenden. Wie vorhergesagt unterschieden sich die Bewertungen der Probanden hinsichtlich der Anwesenheit des ultimativen Effekts, dem Gedanken des letzten Aliens in der Kette, nicht im Hinblick auf die „Reading“-vs.-„Sending“-Manipulation. Die dem Modell zugrunde liegende Idee einer adaptiven ursache-vs. effektseitigen Fehlerattribution konnte damit auch in einer kausalen Struktur bestätigt werden, für die das Modell nicht entwickelt worden war.

In Abschnitt 5.3 wurde schließlich getestet, welchen Einfluss es hat, wenn in eine *Common-Cause*-Struktur weitere beobachtbare präventive Ursachen eingeführt werden, die mit der modellendogenen Fehlerquelle bei der Fehlerattribution konkurrieren. Das Modell sagt in diesem Fall vorher, dass die Einschätzung bezüglich der Anwesenheit eines unbeobachteten Zieleffekts deutlich weniger von der Anzahl der weiteren als anwesend beobachteten Effekte abhängen sollte, wenn in das kausale System eine als anwesend beobachtete präventive Ursache eingeführt wird, die nicht auf den Zieleffekt, aber auf alle

anderen beobachteten Effekte wirkt. In diesem Fall dient die beobachtbare präventive Ursache als Erklärung für die mögliche Abwesenheit von Effekten und reduziert damit die Attribution auf die modellendogen angenommene unbeobachtete Fehlerquelle. In einem ersten Experiment, das an die Basisexperimente angelehnt war, wurden für den „Sending“-Fall drei experimentelle Bedingungen bezüglich einer solchen präventiven Ursache konstruiert. Neben zwei Kontrollbedingungen ohne eine solche Ursache wurden in einer weiteren Bedingung Wände zwischen dem Gedanken sendenden Alien und den beobachteten Effektaliens eingefügt. In dieser Bedingung war der Einfluss der Anzahl der als anwesend beobachteten weiteren Effekte auf die Zielinferenz zwar etwas geringer als in den beiden Kontrollbedingungen, dieser Unterschied konnte statistisch jedoch nicht abgesichert werden, was vermutlich vor allem an der mangelnden expliziten Instruktion in Verbindung mit der interindividuellen Manipulation gelegen haben mag. In einem weiteren Experiment wurde nur für den Fall der Anwesenheit der gemeinsamen Ursache die Anwesenheit einer präventiven Ursache („Wände“ und „Offene Türen“) intraindividuell und damit in der Bedeutung etwas offensichtlicher manipuliert. Wie vorhergesagt zeigte sich eine deutlich geringere Abhängigkeit der Einschätzung der Anwesenheit eines unbeobachteten Zieleffekts von der Anzahl der als anwesend beobachteten Effekte für die Testfälle mit anwesender präventiver Ursache („Wände“). Das Modell hat sich damit auch in einem Kontext bewährt, in dem in die betrachtete *Common-Cause*-Struktur explizit eine weitere präventive Ursache eingeführt wurde, die direkt mit der modellendogenen Fehlerquelle im Hinblick auf die Fehlerattribution konkurriert.

Insgesamt konnte das Modell bzw. die dem Modell zugrunde liegende Idee einer adaptiven Fehlerattribution somit in sehr verschiedenen Kontexten bestätigt werden. Die gute Übereinstimmung der Modellvorhersagen mit den Befunden der Basisexperimente kann also im Hinblick auf die Ausführungen in Abschnitt 4.6 als überzeugend gelten. Sie ist offenkundig weder Ergebnis zu hoher Modellflexibilität – wie in Abschnitt 4.5 dargelegt –, als sie auch nicht schlichtes Ergebnis zu hoher Modellierungsflexibilität sein kann, wie durch die gute Bestätigung in den expliziten Tests im vorliegenden Kapitel belegt wurde.

## 6 Modellerweiterung: Merkmalsbasierte Inferenz

Das in diesem Kapitel dargestellte erweiterte Modell wie auch die empirischen Befunde finden sich in gekürzter Form bereits in Mayrhofer et al. (2008).

### 6.1 Idee und Überblick

In den vorausgegangenen Kapiteln wurde ein Modell entwickelt, das auf der Basis von Annahmen über die Fehlerstruktur in einem kausalen System die Zielinferenzen als adaptive Fehlerattribution modelliert. Ausgangspunkt hierfür waren mögliche Annahmen bzw. Vorwissen der Versuchspersonen über die zugrunde liegenden kausalen Prozesse, wie z.B. Gedankenlesen und Gedankensenden. Welche Mechanismen zugrunde lagen, wurde explizit instruiert, in den meisten der Experimente auch die entsprechenden Konsequenzen bezüglich der Fehlerstruktur. Es erscheint plausibel, dass Menschen auch sensitiv sind für weitere Informationen – insbesondere Merkmale der involvierten Objekte –, wenn sie Vorhersagen über das Auftreten von Ereignissen machen. Dazu ein einfaches Beispiel: Angenommen, man ist zu Hause und beobachtet durch seine Fenster, dass auf dem Fernseher des linken Nachbarn weißes Rauschen läuft, irgendetwas also möglicherweise mit dem Empfang nicht zu stimmen scheint (gebe es zur Vereinfachung nur ein einziges Fernsehprogramm). Auf dem Fernseher des rechten Nachbarn sieht man aber ein Bild. Es stellt sich nun die Frage, ob auf dem eigenen Fernseher ebenfalls ein Bild zu sehen sein wird; schließlich könnte es sein, dass der rechte Nachbar ein Video schaut. Insoweit entspricht das Szenario dem Szenario, das im Basismodell abgebildet ist und eine Fehlerattribution verlangt (ist es ein Problem des Fernsehers des linken Nachbarn oder ein allgemeines Problem). Nehmen wir weiter an, man beobachtet, dass auch der Fernseher des Nachbarn gegenüber weißes Rauschen zeigt, der Fernseher des Nachbarn nach hinten raus aber auch ein Bild. Geht man davon aus, dass Empfangsfehler in einem Fernseher eher sehr selten sind, es also äußerst unwahrscheinlich ist, dass zwei von vier Fernsehern unabhängig voneinander gerade einen Defekt haben, wird man in diesem Fall geneigt sein, auf ein allgemeines Problem zu schließen, und folgern, dass auch der eigene Fernseher nur Rauschen zeigen wird. Hat man nun aber zusätzliche Informationen, z.B. dass die beiden Nachbarn, deren Fernseher ausgefallen sind, einen Kabelanschluss besitzen, die beiden anderen Nach-

barn wie man selbst jedoch eine Satellitenschüssel, dann wird man eher geneigt sein, zu folgern, dass der eigene Fernseher sehr wohl ein Bild zeigen wird – dies auch ohne dass man genaue Vorstellungen von der Funktionsweise der verschiedenen Übertragungssysteme hat.

Im vorliegenden Beispiel scheinen also die Merkmale der involvierten Objekte – hier mithin die Fernseher und die Art ihrer Anschlüsse –, also in gewisser Weise deren Kategorisierung, für die Inferenzen eine bedeutende Rolle zu spielen. Der Einfluss kategorialen Wissens auf Kausalinferenzen wie auch der Einfluss von Kausalwissen auf die Kategorisierung von Objekten wurde dabei bereits vielfach untersucht (siehe u.a. Ahn, 1998; Ahn, Kalish, Gelman, Medin, Luhmann, Atran et al., 2001; Kemp, Goodman & Tenenbaum, 2007; Lien & Cheng, 2000; Liljeholm & Cheng, 2007; Rehder, 2003a, 2003b; Rehder & Burnett, 2005; Rehder & Hastie, 2001; Rehder & Kim, 2006; Waldmann & Hagmayer, 2006). Dabei lassen sich zwei Hauptrichtungen unterscheiden: Zum einen wurde intensiv untersucht, wie das Wissen um den kausalen Status von Merkmalen von Objekten deren Kategorisierung beeinflusst (sogenannte *Causal-Status-Hypothese*, siehe Ahn, 1998; Ahn et al. 2001; Rehder, 2003a, 2003b; Rehder & Kim, 2006). Die Experimente von Rehder und Burnett (2005), die in Abschnitt 2.3 ausführlich beschrieben wurden, sind z.B. aus dieser Perspektive motiviert gewesen. – Zum anderen wurde untersucht, wie Kausallernen die Kategorisierung von Ereignissen/Objekten<sup>48</sup> beeinflusst, insbesondere bezüglich der Gruppierung von Ereignissen/Objekten im Hinblick auf die resultierende Kausalstärkeschätzung („*Causality shapes categories*“; siehe u.a. Lien & Cheng, 2000, die die These vertreten, dass Menschen Kategorien dergestalt bilden, dass sie die kausale Vorhersagekraft maximieren). Auch das Gegenteil wurde bereits gezeigt, also dass vorhandenes kategoriales Wissen in bestimmten Situationen einen Einfluss auf Kausalstärkeschätzungen hat („*Categories shape causality*“; siehe Waldmann & Hagmayer, 2006, sowie Kemp et al., 2007, für einen Ansatz der beide Thesen integriert).

---

<sup>48</sup> Zwischen Ereignissen und Objekten, die diese Ereignisse hervorbringen, wird in der Literatur derzeit kein Unterschied gemacht. In der Gesamtdiskussion in Kapitel 7 soll dieses Problem nochmals aufgegriffen werden, da es dem Autor sinnvoll scheint, zwischen einer statistischen Ebene (Ereignisse, Bayes-Netze) und einer kausalen Hintergrund-Ebene (Objekte mit Merkmalen, Mechanismen etc.) zu unterscheiden.

Nichtsdestotrotz lässt sich obiges Beispiel nicht mit den bereits bekannten Theorien erklären, da die Inferenz auf der Ebene der Kausalstruktur und nicht etwa auf der Ebene der Kausalstärkeschätzung stattfindet. Für das obige Beispiel liegt es daher nahe, dem Basismodell (Kapitel 4) entsprechend für jede Kategorie von Objekten (Fernseher mit Kabelanschluss, Fernseher mit Satellitenanschluss) eine eigene gemeinsame Fehlerquelle zu definieren, die jeweils nur die zu einer Kategorie gehörenden Effekte betrifft. Im obigen Beispiel würde man dann inferieren, dass die zum Kabelanschluss gehörende Fehlerquelle aller Wahrscheinlichkeit nach anwesend ist und damit nicht der eigene Fernseher betroffen ist.

In anderen als den im obigen Beispiel benutzten technischen Kontexten (z.B. sozialen oder natürlichen Kontexten) mag die Inferenz nicht so klar sein, da die kausalen Konsequenzen der beobachteten Merkmale weniger offensichtlich oder mit mehr Unsicherheit behaftet sind und auch mehrere Merkmale in Frage kommen. In diesen Fällen bietet es sich an, über die verschiedenen möglichen Kategorisierungen (oder mehr technisch: Partitionierungen) zu integrieren und diese jeweils mit deren Wahrscheinlichkeit gegeben der beobachteten Merkmale zu gewichten (siehe dazu das *Causal-Schemata*-Modell von Kemp et al., 2007).

Im Folgenden soll daher eine entsprechende Modellerweiterung hergeleitet werden, die für den Fall eines Merkmals die Zielinferenz, also die Einschätzung der Wahrscheinlichkeit der Anwesenheit eines unbeobachteten Zieleffekts gegeben des Status der gemeinsamen Ursache sowie der weiteren Effekte, beschreibt. Aus diesem Modell werden sodann Vorhersagen abgeleitet und in einem Experiment getestet.

## **6.2 Modellierung**

Aus Gründen der Übersichtlichkeit beschränkt sich die Ableitung im Folgenden auf die wesentlichen Schritte. Eine ausführliche Variante findet sich in Anhang K.

### *6.2.1 Ableitung der Zielinferenz*

Ausgangspunkt für die Modellierung ist wie in den vorhergehenden Experimenten wieder die Zielfrage, die die Versuchspersonen zu beantworten hatten, also die Beurteilung der Wahrscheinlichkeit der Anwesenheit eines unbeobachteten

Zieleffekts  $E_n$  in einer *Common-Cause*-Struktur gegeben des Status der gemeinsamen Ursache  $C$  und der anderen beobachteten Effekte  $E_1, \dots, E_{n-1}$ . Ergänzend hinzu treten jedoch Informationen über Merkmale der Effekte. Sei dazu  $F$  eine *Feature*-Matrix, die für jeden der Effekte  $E_1, \dots, E_n$  die Ausprägung eines binären Merkmals (*feature*) dieser Effekte enthält.<sup>49</sup> Dann lässt sich die Zielinferenz schreiben als:

$$P(E_n = 1 | C, E_1, \dots, E_{n-1}, F; \omega, \psi) \quad (23)$$

Der Parametervektor  $\omega$  stellt wieder das abstrakte Vorwissen bzw. die Annahmen der Versuchspersonen über die kausale Domäne dar (siehe dazu Näheres in Abschnitt 4.2) und der Parametervektor  $\psi$  das Wissen um die kategoriale Struktur der Domäne (insbesondere das Wissen um die kausale Relevanz des Merkmals und die diesbezügliche Unsicherheit). Um mit den Parametern der kausalen Domäne rechnen zu können, werden diese analog dem Vorgehen im Basismodell in das System eingeführt und ausintegriert:

$$\begin{aligned} P(E_n = 1 | C, E_1, \dots, E_{n-1}, F; \omega, \psi) \\ = \iint P(E_n = 1 | C, E_1, \dots, E_{n-1}, F; w, \psi) \cdot P(w | \omega) dw \end{aligned} \quad (24)$$

Soweit entspricht das Vorgehen der Ableitung des Basismodells, es stellt sich nun aber die Frage, wie die Inferenz  $P(E_n = 1 | C, E_1, \dots, E_{n-1}, F; w, \psi)$  aufgelöst werden kann. In Anlehnung an das *Causal-Schemata*-Modell von Kemp et al. (2007) können hierbei die in die Inferenz eingehenden Merkmale  $F$  als durch eine generative Kategorienstruktur hervorgebracht gedacht werden. Das heißt, es gibt eine Partitionierung  $z$  der Effekte  $E_1, \dots, E_n$  in höchstens  $n$  Cluster und jeder Cluster  $z_j$  ist mit einer Wahrscheinlichkeit  $\theta_{iz_j}$  verbunden, mit der er das Merkmal  $i$  hervorbringt. Die wahre Partitionierung der Effekte wie auch die Merkmalsstruktur der die Partitionierung bildenden Cluster ist jedoch unbekannt und muss erschlossen werden. Die eigentliche Inferenz bezüglich des Zieleffekts hängt wiederum nur von der Partitionierung, nicht aber von den Merkmalen der Effekte ab. Dazu führen wir die möglichen Partitionierungen in das System ein und integrieren über alle möglichen Partitionierungen  $z$ :

---

<sup>49</sup> Dies dient hier nur der übersichtlicheren Darstellung und kann einfach auf den Fall von  $k$  Merkmalen erweitert werden.



$$\begin{aligned}
& P(E_n = 1 | C, E_1, \dots, E_{n-1}, F; w, \psi) \\
&= \sum_z P(E_n = 1 | C, PN, E_1, \dots, E_{n-1}, z; w) \cdot P(z | F; \psi)
\end{aligned} \tag{25}$$

Auf dieser Ebene handelt es sich also wieder um eine Art zweistufige Inferenz: Gegeben der Merkmale der Effekte kann für jede mögliche Partitionierung eine Wahrscheinlichkeit und gegeben einer Partitionierung kann die auf die Partitionierung bedingte Zielinferenz bestimmt werden. Die letztendliche Zielinferenz, also die Wahrscheinlichkeit der Anwesenheit des Zieleffekts gegeben des Status der Ursache und der anderen Effekte, ergibt sich dann als gewichtetes Mittel der auf die Partitionierungen bedingten Zielinferenzen. Partitionierungen, die aufgrund der beobachteten Merkmale eine hohe Wahrscheinlichkeit haben, haben einen großen Einfluss auf die Zielinferenz, Partitionierungen, die aufgrund der beobachteten Merkmale äußerst unwahrscheinlich sind, haben praktisch keinen Einfluss.

Im Folgenden sollen die beiden Schritte etwas ausführlicher beschrieben und hergeleitet werden.

### 6.2.2 Zielinferenz gegeben einer Partitionierung $z$

Wäre die kausal relevante Partitionierung  $z$  der Effekte bekannt, ließe sich die Zielinferenz relativ einfach bestimmen. Wie oben bereits beschrieben, ist die grundsätzliche Modellierungsidee, dass die Effekte innerhalb eines Clusters  $z_j$  eine gemeinsame Fehlerquelle  $PN_{z_j}$  teilen. Das heißt, es wird davon ausgegangen, dass es nur Abhängigkeiten zwischen Effekten innerhalb eines Clusters gibt, aber nicht zwischen Effekten verschiedener Cluster (für Beispiele siehe Abbildung 36). Dies lässt sich als vereinfachende Annahme dadurch rechtfertigen, dass es sich bei der als bekannt angenommenen Partitionierung um eine kausal relevante Partitionierung handelt<sup>50</sup>. (Das Problem für die Zielinferenz ist aber, dass die relevante Partitionierung bzw. die relevanten Partitionierungen aus den beobacht-

<sup>50</sup> Man könnte argumentieren, dass es neben den cluster-spezifischen gemeinsamen Fehlerquellen noch eine gemeinsame Fehlerquelle geben sollte, die auf alle Effekte unabhängig von ihrer Clusterzugehörigkeit wirkt. Da jedoch aufgrund der Aggregation über die verschiedenen Partitionierungen auch die Partitionierung in die Zielinferenz eingeht, die alle Effekte in einem Cluster vereinigt, mithin also alle Effekte mit einer gemeinsamen Fehlerquelle ausstattet, scheint diese Fehlerquelle auf Partitionierungsebene entbehrlich.

baren Merkmalen erschlossen werden muss, was selbst wiederum mit Unsicherheit behaftet ist.)

Seien nun für eine gegebene Partitionierung  $z$   $E_{z_j}$  diejenigen Effekte, die zum  $j$ ten Cluster innerhalb der Partitionierung gehören ( $E_n$ , der Zieleffekt, selbst nicht eingeschlossen) und  $z_n$  der Cluster, der  $E_n$  enthält, und damit  $E_{z_n}$  diejenigen Effekte, die sich im selben Cluster wie  $E_n$  befinden. Da  $E_n$  nur mit den Effekten  $E_{z_n}$  eine gemeinsame Fehlerquelle,  $PN_{z_n}$ , teilt, nicht aber mit den anderen Effekten der *Common-Cause*-Struktur, lässt sich die Zielinferenz gegeben einer Partitionierung  $z$  analog zu Gleichung 11 in Abschnitt 4.2 vereinfacht schreiben als:

$$\begin{aligned} P(E_n = 1 | C, E_1, \dots, E_{n-1}, z; w) \\ = \sum_{PN_{z_n}} P(E_n = 1 | C, PN_{z_n}; w) \cdot P(PN_{z_n} | C, E_{z_n}; w) \end{aligned} \quad (26)$$

Damit entspricht die Zielinferenz gegeben einer Partitionierung  $z$  exakt dem Basismodell (siehe Kapitel 4), wenn man sich diejenigen Effekte hinwegdenkt, die sich nicht im selben Cluster wie  $E_n$  befinden.

In Abbildung 36 sind zur Veranschaulichung für eine *Common-Cause*-Struktur mit fünf Effekten beispielhaft vier der 52 möglichen Partitionierungen<sup>51</sup> und die zugehörigen Fehlerquellen dargestellt. In Teilabbildung (a) gehören alle Effekte zu einem Cluster, d.h. sie teilen sich auch eine gemeinsame Fehlerquelle und damit hängt die Zielinferenz, z.B. hinsichtlich des Effektes  $E_5$ , von allen anderen Effekten ab. In Teilabbildung (b) und (c) sind die Effekte in zwei Cluster aufgeteilt. Die Zielinferenz hinsichtlich  $E_5$  hängt hier nur von  $E_4$  (b) bzw. von  $E_3$  und  $E_4$  (c) ab. In Teilabbildung (d) gehört jeder Effekt zu einem eigenen Cluster, die Zielinferenz über  $E_5$  ist damit unabhängig von den anderen Effekten; es handelt sich auf der Ebene der beobachtbaren Variablen also um ein Bayes-Netz mit Markov-Bedingung (siehe für eine analoge Herleitung für ein *Causal-Chain*-Modell Abschnitt 5.2.1).

<sup>51</sup> Die Anzahl der möglichen Partitionierungen einer Menge von  $n$  Objekten wird durch die Bellsche Zahl beschrieben.

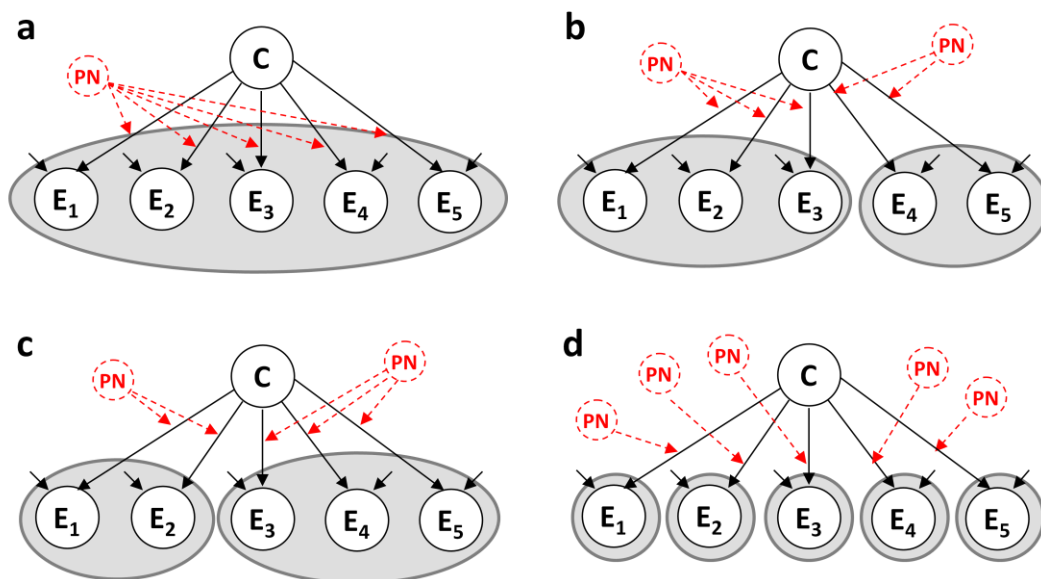


Abbildung 36. Vier Beispiele für ein um eine für jeden Cluster je gemeinsame Fehlerursache  $PN_{z_j}$  erweitertes *Common-Cause-Modell* mit fünf Effekten. In Teilabbildung (a) gehören alle Effekte zum selben Cluster (diese Partitionierung ist mithin äquivalent zum Basismodell), in Teilabbildung (b) und (c) sind die Effekte in jeweils zwei Cluster partitioniert und in Teilabbildung (d) gehört jeder Effekt zu einem eigenen Cluster (diese Partitionierung ist auf der Ebene der beobachteten Variablen mithin äquivalent zu einem Bayes-Netz). Konditionale Abhängigkeiten zwischen Effekten gibt es mithin nur innerhalb von Clustern, aber nicht zwischen diesen.

### 6.2.3 Wahrscheinlichkeit einer Partitionierung $z$ gegeben $F$

Die verbleibende Frage ist nun, mit welchem jeweiligen Gewicht  $P(z|F)$  die eben abgeleiteten, einzelnen bedingten Zielinferenzen zur Gesamtinferenz aggregiert werden (siehe dazu Gleichung 25 auf Seite 122). Dazu muss für jede mögliche Partitionierung die *A-posteriori*-Wahrscheinlichkeit gegeben der beobachteten Merkmale bestimmt werden. Dies erfolgt im ersten Schritt einfach durch Anwendung der Bayes-Formel:

$$P(z|F; \psi) \propto P(F|z; \psi) \cdot P(z) \quad (27)$$

Die Wahrscheinlichkeit einer Partitionierung  $z$  gegeben der Merkmale  $F$  ist demnach proportional zum Likelihood der Merkmale gegeben der Partitionierung gewichtet mit deren *A-priori*-Wahrscheinlichkeit.

Die *A-priori*-Wahrscheinlichkeit  $P(z)$  lässt sich dabei als *Chinese-Restaurant-Prozess (CRP)* beschreiben (siehe Pitman, 2002; Kemp et al., 2007).

Ein *CRP* ordnet jeder Partitionierung  $z$  von  $n$  Objekten eine Wahrscheinlichkeit dahingehend zu<sup>52</sup>, dass Partitionierungen, die nur wenige, große Cluster enthalten, deutlich wahrscheinlicher sind, als Partitionierungen, die aus vielen kleinen Clustern bestehen. Dies soll eine *A-priori*-Tendenz ausdrücken, die Objekte in der Umwelt möglichst sparsam, also in wenigen Kategorien zu organisieren.

Für die Partitionierungen der Effekte in Abbildung 36 (siehe auch Abbildung 37) gilt damit folgende *A-priori*-Tendenz – also sofern die Merkmale ignoriert werden:  $P(z^a) > P(z^b) = P(z^c) > P(z^d)$ . Das heißt, ohne Beachtung der Merkmale ist die Partitionierung, die nur einen großen Cluster enthält, aufgrund des *CRP*-Priors viel wahrscheinlicher, als die Partitionierung, die fünf kleine Cluster enthält. Die Partitionierungen in den Teilabbildungen (b) und (c) sind *a priori* gleich wahrscheinlich, da sie je zwei Cluster gleicher Größe enthalten.

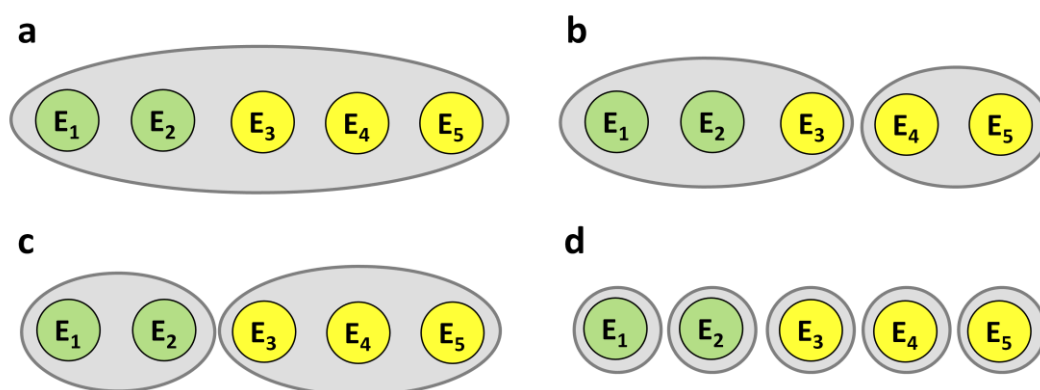


Abbildung 37. Die vier in Abbildung 36 dargestellten Beispiele für mögliche Partitionierungen von fünf Effekten ergänzt mit einer binären Eigenschaft (gelb vs. grün) zur Illustration des Einflusses von Eigenschaftsausprägungen auf die Wahrscheinlichkeit einer bestimmten Partitionierung. Die Effekte  $E_1$  und  $E_2$  sind grün, die Effekte  $E_3$ ,  $E_4$  und  $E_5$  gelb.

Das Likelihood einer Merkmalskonfiguration  $F$  gegeben einer bestimmten Partitionierung  $z$  lässt sich unter der vereinfachten Annahme nur eines binären Merkmals (mit den Ausprägungen A und B) leicht als Beta-Binomial-Modell beschreiben, in welchem jedem Cluster  $z_j$  eine Wahrscheinlichkeit  $\theta_{z_j}$  zugeordnet wird, mit der dieser Cluster die Merkmalsausprägung A hervorbringt (also z.B.

<sup>52</sup> Es gibt auch *CRP*-Varianten, die freie Parameter haben. Diese wurden derzeit in den in der Literatur bereits verwendeten Modellen jedoch nicht eingesetzt; daher soll auch hier aus Gründen der Einfachheit darauf verzichtet werden. Der *CRP* ist verwandt zur Ewens-Verteilung, die u.a. in der Biodiversitäts-Forschung eingesetzt wird.

gelb zu sein). Entsprechend beschreibt die Gegenwahrscheinlichkeit  $1 - \theta_{z_j}$  die Wahrscheinlichkeit, mit der der Cluster die Merkmalsausprägung B, also z.B. grün zu sein, hervorbringt. Da die Wahrscheinlichkeit  $\theta_{z_j}$  unbekannt ist, definieren wir für diese eine *A-priori*-Verteilung, und zwar unter der Annahme, dass Cluster von den Merkmalsausprägungen her homogen sind,  $\theta_{z_j}$  also entweder nahe 0 oder nahe 1 liegt (hierfür bietet sich eine Beta-Verteilung mit sehr kleinen Parametern an)<sup>53</sup>. Seien nun  $F_{z_j}$  die Merkmalsausprägungen der Effekte, die dem Cluster  $z_j$  zugeordnet sind. Dann ergibt sich unter der Annahme, dass die Cluster hinsichtlich der Merkmale unabhängig voneinander sind und sich damit das Gesamtl likelihood als Produkt der Likelihoods über die Cluster darstellt:

$$P(F|z; \psi) = \prod_{z_j} P(F_{z_j}|\psi) = \prod_{z_j} \int P(F_{z_j}|\theta_{z_j}) \cdot P(\theta_{z_j}; \psi) d\theta_{z_j} \quad (28)$$

mit

$$P(F_{z_j}|\theta_{z_j}) \propto (\theta_{z_j})^{\#(F_{z_j}=A)} \cdot (1 - \theta_{z_j})^{\#(F_{z_j}=B)} \quad (29)$$

als Binomial-Modell und

$$P(\theta_{z_j}; \psi) \sim \text{Beta}(\alpha_\theta, \beta_\theta) \quad (30)$$

als Beta-Prior.

Innerhalb eines Clusters  $z_j$  bestimmt sich damit das Likelihood des beobachteten Merkmals entsprechend der Binomial-Verteilung: Ist – wie oben beschrieben – die Wahrscheinlichkeit der Merkmalsausprägung A gleich  $\theta_{z_j}$  und entsprechend die Wahrscheinlichkeit der Merkmalsausprägung B gleich  $(1 - \theta_{z_j})$ , dann ist das Likelihood genau  $[\#(F_{z_j} = A)]$ -mal die Ausprägung A und  $[\#(F_{z_j} = B)]$ -mal die Ausprägung B zu beobachten, proportional zu Gleichung 29.

Betrachtet man diesbezüglich nun Abbildung 37, dann sieht man, dass die in den Teilabbildungen (c) und (d) dargestellten Partitionierungen im Hinblick auf das betrachtete Merkmal der Effekte eine sehr hohe Wahrscheinlichkeit haben, da alle enthaltenen Cluster perfekt homogen sind, d.h. die zu wählenden  $\theta$ s ent-

---

<sup>53</sup> Zum Beispiel eine  $\text{Beta}(\frac{1}{25}, \frac{1}{25})$ -Verteilung, die umgekehrt u-förmig ist und viel Wahrscheinlichkeitsgewicht nahe 0 und nahe 1 hat.

weder 0 oder 1 sind. Die in den Teilabbildungen (a) und (b) dargestellten Partitionierungen haben jedoch ein deutlich geringeres Likelihood, da die Cluster nicht bzw. zum Teil nicht homogen sind.

Nimmt man nun die *A-Priori*-Wahrscheinlichkeit einer Partitionierung und deren Likelihood zusammen, sieht man z.B. für die in Abbildung 37c dargestellte Partitionierung, dass diese wahrscheinlich das höchste Gewicht erhalten wird. Zwar haben Partitionierung (c) und (d) das gleiche Likelihood, (c) ist aber *a priori* viel wahrscheinlicher. Genauso sind (b) und (c) *a priori* gleich wahrscheinlich, (c) hat aber das höhere Likelihood.

#### 6.2.4 Modellvorhersagen

Die Modellvorhersage selbst ergibt sich wie in Abschnitt 6.2.1 beschrieben als gewichteter Mittelwert aller Zielinferenzen über alle möglichen Partitionierungen. Welche Vorhersage ergibt sich nun für die Wahrscheinlichkeit der Anwesenheit des Zieleffekts gegeben des Status der Ursache, des Status der weiteren Effekte und deren Features? Betrachten wir hierfür eine *Common-Cause*-Struktur mit fünf Effekten, die je eine binäre Eigenschaft haben, wie sie in Abbildung 36 und Abbildung 37 dargestellt ist. In einer solchen Struktur kann nun nach den Herleitungen in den vorangehenden Abschnitten qualitativ erwartet werden, dass die Inferenz über die Anwesenheit eines unbeobachteten Zieleffekts, z.B.  $E_5$ , umso stärker vom Status eines anderen Effekts abhängig ist, je ähnlicher dieser dem Zieleffekt ist. Eine Einschätzung bezüglich der Anwesenheit des Effekts  $E_5$  sollte mithin also – entsprechende Annahmen über die zugrundeliegenden Kausalprozesse vorausgesetzt – stark vom Status der Effekte  $E_3$  und  $E_4$ , aber weniger stark – weil eine andere Merkmalsausprägung aufweisend – von den Effekten  $E_1$  und  $E_2$  abhängig sein.

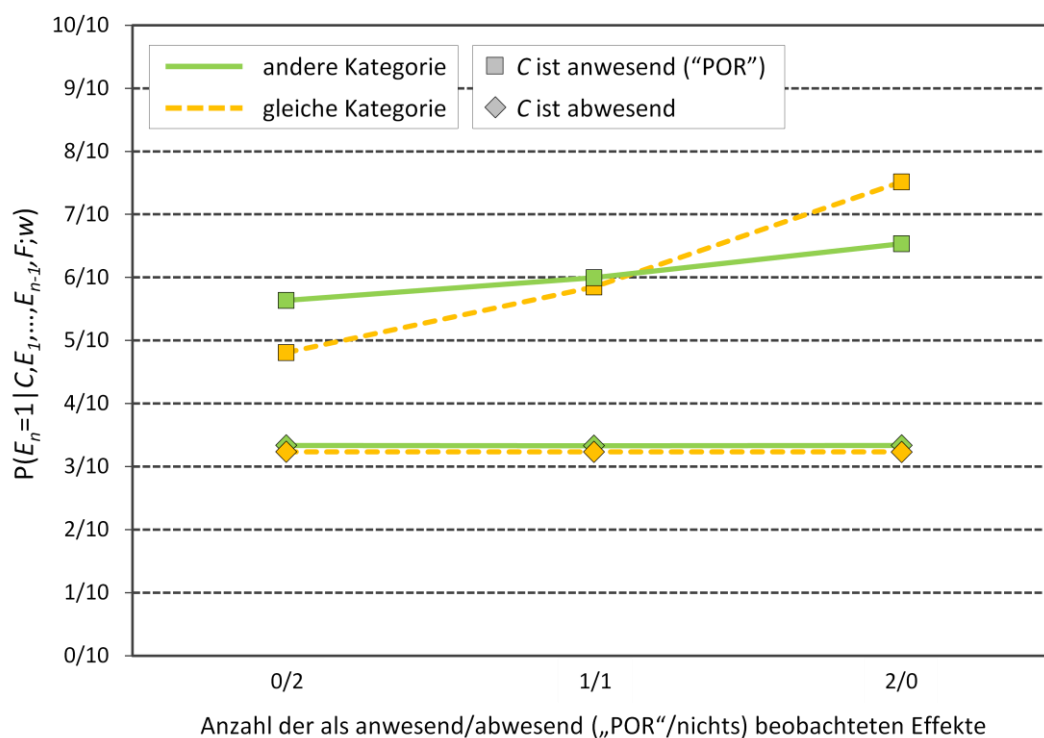
Für eine konkrete Vorhersage müssen wieder Annahmen über die zugrundeliegenden Parameter getroffen werden. Für die kausalen Zielinferenzen, wie sie für eine beliebig gegebene Partitionierung  $z$  in Abschnitt 6.2.2 abgeleitet wurden, können einfach die Parameterverteilungen verwendet werden, wie sie bereits für die Vorhersagen des Basismodells in Abschnitt 4.3 genutzt wurden. Dabei soll sich im Folgenden allerdings auf die Entsprechung zur „*Sending*“-Bedingung beschränkt werden, also ein Fall mit einer bzw. bei mehreren Clustern

entsprechend mit mehreren starken gemeinsamen Fehlerquellen [ $w_{PN} \sim \text{Beta}(10,1)$ ]. Dies erscheint deshalb sinnvoll, weil eine starke Abhängigkeit vom Status ähnlicher Effekte ja voraussetzt, dass die Annahmen über die zugrunde liegenden Kausalprozesse eine solche starke Abhängigkeit implizieren, wie dies in der „*Sending*“-Bedingung der Fall war, denn nur dann kann auch eine substantielle Abnahme dieser Abhängigkeit mit steigender Unähnlichkeit der Effekte erwartet werden.

Für die *A-priori*-Verteilung der möglichen Partitionierungen kommt wie oben beschrieben ein *Chinese-Restaurant*-Prozess zum Einsatz (Tendenz zu wenigen großen Clustern); die cluster-spezifischen Merkmalswahrscheinlichkeiten  $\theta_{z_j}$  werden aus einer  $\text{Beta}(1/25, 1/25)$  gezogen (Tendenz zu homogenen Clustern bezüglich der Verteilung des Features).

In Abbildung 38 ist die Vorhersage des Modells dargestellt, wie sie aus einer Monte-Carlo-Simulation mit 100 000 Durchgängen gewonnen wurde. Zur besseren Darstellung sind die Vorhersagen über die Zustände der Effekte der jeweils anderen Kategorie gemittelt abgetragen. (Die ungemittelten Vorhersagen finden sich in Anhang L.)

Die gestrichelte, gelbe Linie stellt dabei die Zielinferenz über den Zieleffekt in Abhängigkeit von der Anzahl der als anwesend beobachteten Effekte mit der gleichen Farbe dar (der Zieleffekt sei einmal ebenfalls gelb). Man sieht, dass die Wahrscheinlichkeit deutlich mit steigender Anzahl ansteigt, wenn die gemeinsame Ursache *C* anwesend ist. Die durchgezogene, grüne Linie zeigt die Abhängigkeit der Zielinferenz von der Anzahl der als anwesend beobachteten Effekte der anderen Kategorie (grün). Deutlich zu erkennen ist – wie bereits qualitativ vorhergesagt – eine merklich geringere Abhängigkeit.



**Abbildung 38.** Modellvorhersage auf der Basis einer Monte-Carlo-Simulation mit 100 000 Durchgängen in einem *Common-Cause*-Modell mit einer Ursache  $C$  und fünf Effekten. Die Parameter des kausalen Systems wurden hierfür aus den gleichen Verteilungen gezogen, die für die „*Sending*“-Bedingung im Basismodell Verwendung fanden. Der Parameter für die cluster-spezifischen Merkmalswahrscheinlichkeiten wurde aus einer  $Beta(1/25, 1/25)$ -Verteilung gezogen. Die *A-priori*-Wahrscheinlichkeiten der möglichen Partitionen entstammen einem *Chinese-Restaurant-Prozess (CRP)*.

### 6.3 Experiment 8

Die Vorhersage des Modells, dass der Einfluss des Status der Effektvariablen von deren weiteren (kausal relevanten) Merkmalen abhängt, soll im Folgenden empirisch im bereits in den vorhergehenden Experimenten benutzten Paradigma getestet werden. Hierfür bietet es sich an, für den „*Sending*“-Fall – der ja starke Abhängigkeiten repräsentiert – Effektliaien mit unterschiedlichen Merkmalen zu benutzen. Das Modell sagt vorher, dass im Hinblick auf die Bewertung der Anwesenheit eines Zieleffekts der Einfluss derjenigen Effekte, die andere Merkmalsausprägungen haben als der Zieleffekt, deutlich geringer ausfallen sollte, als der Einfluss derjenigen Effekte, die die gleichen Eigenschaften haben wie der Zieleffekt. Als saliente Eigenschaft kann hierfür die Farbe der Aliens herangezogen



werden, die für unterschiedliche Alien-Arten stehen soll, um eine kausale Relevanz nahe zu legen.

Das experimentelle Basisdesign lässt sich nun leicht dahingehend erweitern, dass dem Zielalien nicht nur zwei Effekte mit der gleichen Farbe (also der gleichen Alien-Art) zur Seite gestellt werden, sondern zusätzlich noch zwei weitere Effektlialiens mit einer anderen Farbe (also einer anderen Alien-Art). In diesem erweiterten *Common-Cause-Modell* mit nun fünf Effekten lässt sich separat testen, in welchem Ausmaß die Gedanken der Effektlialiens gleicher Kategorie und in welchem Ausmaß die Gedanken der Effektlialiens anderer Kategorie die Beurteilungen hinsichtlich der Anwesenheit des Zieleffekts beeinflussen.

### 6.3.1 Methode

*Versuchspersonen.* An der Untersuchung nahmen 60 Personen teil (33 Frauen, 27 Männer, Durchschnittsalter 25.1 Jahre, Altersspanne 19–56 Jahre). Diese wurden auf dem Campus, im Zentralen Hörsaalgebäude (ZHG), im Psychologischen Institut, in den Mensen rund um den Campus der Universität Göttingen und zum Teil im Bekannten- und Verwandtenkreis der Versuchsleiterinnen und Versuchsleiter angeworben. Es handelte sich dabei größtenteils um Studierende der Universität Göttingen verschiedenster Fachrichtungen. Keiner der Probanden hatte bereits an einem der vorhergehenden Experimente der Serie teilgenommen. Die Teilnahme erfolgte freiwillig. Nach Ende der Untersuchung wurden die Teilnehmerinnen und Teilnehmer mit einer Süßigkeit belohnt und sofern gewünscht über den Zweck und das Ziel der Untersuchung genauer aufgeklärt.

*Versuchsdurchführung und Material.* Experiment 8 bestand wie bereits die vorhergehenden Experimente aus einer Instruktions- und einer Testphase. In der *Instruktionsphase* wurde den Probanden eine vierseitige Instruktion vorgelegt. In dieser wurde den Versuchspersonen erklärt, dass es sechs Aliens – Gonz, Murks, Mezukt, Zoohng, Brxxx und Karkas – gebe, die meistens an nichts und manchmal an „POR“ denken würden. Des Weiteren wurde den Probanden mitgeteilt, dass Gonz seine Gedanken aussenden und damit in die Köpfe der anderen Aliens übertragen könne. Im Konkreten hieße dies, dass die Aliens häufig dann an „POR“ dächten, wenn Gonz an „POR“ denke. Allerdings sei Gonz nicht perfekt, so

dass er es nicht immer schaffe, seine Gedanken auch auszusenden und damit in die Köpfe der anderen Aliens zu übertragen. In einem solchen Fall dächten die Aliens jeder für sich genommen manchmal an „POR“, unabhängig davon, was Gonz gerade denke. Überdies habe sich gezeigt, dass Mezukt und Zoohng zu einer anderen Alien-Art gehörten als Karkas und Brxxx. – Die Alien-Art wurde im Bildmaterial durch die Farbe gekennzeichnet: grün und gelb (siehe Abbildung 39); für die Effektliaens wurde dabei auf dieselbe Abbildung zurückgegriffen (die lediglich unterschiedlich eingefärbt wurde), um den Einfluss impliziter Merkmale auszuschließen. Einen Hinweis auf die kausale Relevanz des Merkmals und darauf, was die Alien-Art mit der Gedankenübertragung zu tun haben könnte, wurde nicht gegeben.

An der Relation Gonz–Murks wurden wieder die vier möglichen Fälle erklärt. Der Text entsprach dem Wortlaut für die „*Sending*“-Bedingung aus Experiment 5.

Vor Abschluss der Instruktionsphase sollten die Probanden die Instruktion in ihren wesentlichen Inhalten noch einmal wiedergeben und diese bei Missverständnissen nochmals lesen.

In der *Testphase* wurden den Probanden wieder sechs Testfälle vorgelegt. Die Effektliaens waren hierfür in einem Halbkreis unterhalb von Gonz angeordnet (siehe dazu den Beispiel-Testfall in Abbildung 39); das Zielalien, immer Murks, direkt unterhalb von Gonz. Die unterschiedlichen Alien-Arten wurden durch verschiedene Farben (gelb und grün) und verschiedene Positionen gekennzeichnet. Die beiden linken Aliens, Mezukt und Zoohng, waren grün und die beiden rechten Aliens, Brxxx und Karkas waren gelb. Die Farbe von Murks, und damit zu welcher Alien-Art er gehört, war über die Versuchspersonen ausbalanciert. In jedem Testfall gehörten damit immer zwei Effektliaens zur gleichen Kategorie wie das Zielalien und zwei Effektliaens zur anderen Kategorie. Für jeden vorgelegten Testfall, also einer bestimmten Konfiguration von Gedanken der diesmal sechs Aliens, hatten die Versuchspersonen anzugeben, in wie vielen von zehn vorgestellten Situationen das Zielalien, Murks, an „POR“ denken würde.

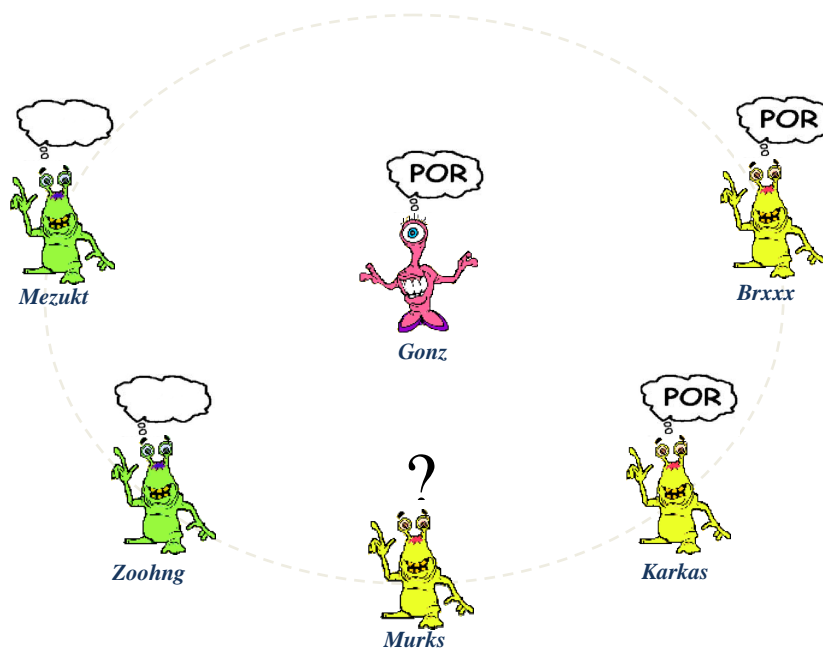


Abbildung 39. Ein Beispiel-Testfall, wie er in der Testphase von Experiment 8 benutzt wurde (verkleinerte Darstellung). Die Versuchspersonen hatten sich zehn Situationen mit der gegebenen Konfiguration vorzustellen und dann anzugeben, in wie vielen dieser Situationen das Zielalien, Murks, an „POR“ denken würde (Skala von 0 bis 10). Mezukt und Zoohng, mit grüner Farbe gekennzeichnet, gehören zu einer anderen Alien-Art als die mit gelber Farbe gekennzeichneten Aliens Murks, Karkas und Brxxx.

Für die Versuchsdurchführung wurde ein Aktenordner angelegt, der die Instruktionen (je vier Seiten für die „Murks-ist-grün“-vs.-„Murks-ist-gelb“-Balancierung), die 64 möglichen Testfälle<sup>54</sup> sowie die Datentabelle und eine mit Excel erzeugte Randomisierungstabelle enthielt. Die Durchführung erfolgte auf dem Campus der Universität Göttingen und in umliegenden Gebäuden. Sie dauerte ca. 10–15 Minuten pro Versuchsperson.

*Design.* Betrachtet man neben dem Zustand der gemeinsamen Ursache (Faktor C; zweistufig: anwesend vs. abwesend) die Anzahl der als anwesend beobachteten Effekte der gleichen Kategorie (Faktor  $N_{E+}^-$ ; dreistufig: 0, 1 oder 2) sowie die Anzahl der als anwesend beobachteten Effekte der anderen Kategorie (Faktor  $N_{E+}^+$ ; dreistufig: 0, 1 oder 2) als getrennte Faktoren<sup>55</sup>, so ergibt sich ein 2x3x3-

<sup>54</sup> Für jede der beiden Balancierungsbedingungen  $2^5 = 32$  Fälle, da fünf Aliens an nichts oder an „POR“ denken; insgesamt also 64 Testfälle.

<sup>55</sup> Die sich jeweils für die Faktorstufe 1 (ein Effekalien denkt an „POR“) ergebenden Wahlmöglichkeiten bei den Testfällen wurde wieder durch zufällige Auswahl gelöst.

Design mit drei Messwiederholungsfaktoren. Da sich daraus in vollständiger Erfassung 18 Testfälle ergeben würden, wurden die Versuchspersonen nunmehr je Ausprägung des Faktors  $C$  (Gonz denkt an nichts vs. Gonz denkt an „POR“) in einem balancierten, partiell-konfundierten randomisierten Blockdesign (siehe Kirk, 1995) dergestalt getestet, dass von den neun Kombinationsmöglichkeiten der Faktoren  $N_{E+}^{\bar{=}}$  und  $N_{E+}^{\neq}$  ( $3 \times 3$ ) jeweils nur drei vorgelegt wurden, so dass jede Ausprägung jedes Faktors genau einmal realisiert wurde (jeweils drei Versuchspersonen konstituieren damit einen vollständigen Datensatz).

*Vorhersage.* Die empirische Vorhersage ergibt sich aus der Vorhersage des auf den Kategorienfall erweiterten Basismodells (siehe Abschnitt 6.2). Wenn die gemeinsame Ursache  $C$  anwesend ist, sollten im Durchschnitt deutlich höhere Bewertungen der Anwesenheit des Zieleffekts auftreten, als wenn diese abwesend ist (Haupteffekt  $C$ ). Für den Fall der Abwesenheit der Ursache ( $C=0$ ) ergibt sich wie im Basismodell die Vorhersage, dass es keinen Einfluss der Anzahl der als anwesend beobachteten Effekte sowohl der gleichen wie auch der anderen Kategorie geben sollte (keine Haupteffekte für  $C=0$  bezüglich  $N_{E+}^{\bar{=}}$  und  $N_{E+}^{\neq}$  wie auch keinen Interaktionseffekt  $N_{E+}^{\bar{=}} \times N_{E+}^{\neq}$  für  $C=0$ ). Ist die gemeinsame Ursache jedoch anwesend ( $C=1$ ), so wird erwartet, dass es einen deutlichen Einfluss der Anzahl der als anwesend beobachteten Effekte sowohl der gleichen wie auch der anderen Kategorie gibt (Haupteffekte für  $C=1$  bezüglich  $N_{E+}^{\bar{=}}$  und  $N_{E+}^{\neq}$ ). Eine Interaktion zwischen den beiden Faktoren – also die Frage, ob die Anzahl der als anwesend beobachteten Effekte der gleichen Kategorie im Hinblick auf die Anzahl der als anwesend beobachteten Effekte der anderen Kategorie unterschiedlich wirkt – sollte unbedeutend ausfallen (siehe dazu Abbildung L-1 und Abbildung L-2 in Anhang L).

Des Weiteren wird erwartet – als wichtigste Modellvorhersage –, dass die Anzahl der als anwesend beobachteten Effekte der gleichen Kategorie wie der Zieleffekt einen deutlich stärkeren Einfluss hat als die Anzahl der als anwesend beobachteten Effekte der anderen Kategorie (mithin Effektgröße  $[N_{E+}^{\bar{=}}] > \text{Effektgröße } [N_{E+}^{\neq}]$  für  $C=1$ ).

### 6.3.2 Ergebnisse

In Tabelle 10 und Abbildung 40 sind die Ergebnisse von Experiment 8 dargestellt. Abbildung 40 zeigt dabei getrennt für beide Stufen des Faktors  $C$  marginalisierte Haupteffekte bezüglich der Faktoren  $N_{E+}^-$  und  $N_{E+}^\neq$ , d.h. es sind jeweils die über die Stufen des anderen Faktors aggregierten Bewertungen abgetragen<sup>56</sup>. Die detaillierten varianzanalytischen Ergebnisse finden sich in Tabelle N-1 und Tabelle N-2 in Anhang N; die nicht-aggregierten Daten sind ergänzend in Abbildung M-1 und Abbildung M-2 in Anhang M dargestellt.

Wie bereits in den vorhergehenden Experimenten ist die Bewertung der Wahrscheinlichkeit der Anwesenheit des Zieleffekts deutlich höher, wenn die Ursache anwesend als wenn diese abwesend ist<sup>57</sup>.

Im Fall der Abwesenheit der Ursache (Gonz denkt an nichts; die beiden unteren Linien in Abbildung 40) steigen die Beurteilungen bezüglich der Anwesenheit des Zieleffekts sowohl mit der Anzahl der als anwesend beobachteten Effekte der gleichen Kategorie von 2.35 auf 2.90 nur unwesentlich und wie vorhergesagt nicht signifikant an (Haupteffekt  $N_{E+}^-$  für  $C=0$ :  $F_{2,112}=1.91$ ,  $p=.15$ ); noch geringer steigen sie mit der Anzahl der als anwesend beobachteten Effekte der anderen Kategorie von 2.50 auf 2.78 (Haupteffekt  $N_{E+}^\neq$  für  $C=0$ :  $F_{2,112}<1$ ,  $p=.51$ ). Die Interaktion ist nicht signifikant (Interaktionseffekt  $N_{E+}^- \times N_{E+}^\neq$ , für  $C=0$ :  $F_{4,112}=1.00$ ,  $p=.41$ ), wie ebenfalls erwartet.

Ist die Ursache anwesend (Gonz denkt an „POR“; die beiden oberen Linien in Abbildung 40) steigen die durchschnittlichen Bewertungen im Hinblick auf die Anzahl der als anwesend beobachteten Effekte der gleichen Kategorie deutlich von 5.52 auf 7.82 an (Haupteffekt  $N_{E+}^-$  für  $C=1$ :  $F_{2,112}=22.53$ ,  $p<.01$ ,  $\eta^2=.26$ ); ebenso, aber deutlich schwächer steigen die Bewertungen mit der Anzahl der als anwesend beobachteten Effekte der anderen Kategorie von 5.87 auf 6.85 an

<sup>56</sup> Diese Darstellungsform wurde gewählt, weil sie den jeweiligen Einfluss (Haupteffekt) der Anzahl der als anwesend beobachteten Effekte der gleichen Kategorie (Faktor  $N_{E+}^-$ ) wie auch der anderen Kategorie ( $N_{E+}^\neq$ ) einfach darstellt und damit eine „grafische“ Beurteilung der Frage, welcher Faktor stärker wirkt, überhaupt erst ermöglicht.

<sup>57</sup> Eine genaue varianzanalytische Prüfung dieser Aussage ist nicht möglich, da der Autor es versäumt hat, das komplexe varianzanalytische Design für die beiden Stufen des Faktors  $C$  dahingehend abzustimmen, dass eine formal korrekte Auswertung über diesen Faktor möglich wäre. Der Fokus beim Design des Experiments lag jedoch bei der Prüfung der Hypothesen, die für jede Stufe des Faktors  $C$  separat formuliert waren. Schätzt man das zugehörige  $p$  für den Haupteffekt  $C$  konservativ, indem man das partiell-konfundierte Blockdesign als interindividuelle Manipulation interpretiert, ergibt sich ein Schätzer für  $p$  der unter  $10^{-5}$  liegt.

(Haupteffekt  $N_{E+}^{\neq}$  für  $C=1$ :  $F_{2,112}=6.60$ ,  $p<.01$ ,  $\eta^2=.08$ ). Die Interaktion ist – wie erwartet – nur unwesentlich ausgeprägt und nicht signifikant (Interaktionseffekt  $N_{E+}^{\neq} \times N_{E+}^{\neq}$ , für  $C=1$ :  $F_{4,112}=1.44$ ,  $p=.27$ ).

Zur Prüfung der Haupthypothese, also der Frage, ob die Anzahl der als anwesend beobachteten Effekte der gleichen Kategorie (Faktor  $N_{E+}^{\neq}$ ) einen größeren Einfluss auf die Bewertungen hat als die Anzahl der als anwesend beobachteten Effekte der anderen Kategorie (Faktor  $N_{E+}^{\neq}$ ), wurde ein Bootstrap-Test konstruiert, da im verwendeten Design keine Standard-Prozedur zur Verfügung steht<sup>58</sup>. Die interessierende Frage ist also, ob der „Gleiche-Kategorie“-Faktor ( $N_{E+}^{\neq}$ ) im Hinblick auf die Bewertungen mehr Varianz erklärt als der „Andere-Kategorie“-Faktor ( $N_{E+}^{\neq}$ ). Eine angemessene Statistik, um dies zu testen, ist das Verhältnis der durch die Faktoren erklärten Varianzen, welches 3.413 beträgt. Um dieses Verhältnis auf Signifikanz zu prüfen, wurde eine Bootstrap-Verteilung für die Nullhypothese, dass die Varianzen gleich sind (d.h. das Verhältnis ist 1), konstruiert, indem 10 000 Bootstrap-Stichproben aus den Daten gezogen wurden unter der Annahme, dass die beiden Faktoren austauschbar sind (d.h. die Kategorie eines Effekts hat keinen Einfluss). Der  $p$ -Wert des empirischen Varianzverhältnisses beträgt 0.026. Demnach hat die Anzahl der als anwesend beobachteten Effekte der gleichen Kategorie auf die Wahrscheinlichkeitsbewertungen einen signifikant größeren Einfluss als die Anzahl der als anwesend beobachteten Effekte der anderen Kategorie. Dies bestätigt die aus dem Modell abgeleitete Hauptvorhersage.

---

<sup>58</sup> Zumindest nach Kenntnis des Autors.

Tabelle 10

*Durchschnittliche Bewertung der Anwesenheit des Zieleffekts (Experiment 8)*

		$N_{E+}^{\bar{=}} = 0$	$N_{E+}^{\bar{=}} = 1$	$N_{E+}^{\bar{=}} = 2$	$\emptyset N_{E+}^{\bar{=}}$
C = 1 („POR“)	$N_{E+}^{\neq} = 0$	4.95 (2.65)	5.20 (2.69)	7.45 (1.32)	5.87 (2.53)
	$N_{E+}^{\neq} = 1$	5.95 (2.56)	7.10 (1.52)	8.05 (1.23)	7.03 (2.02)
	$N_{E+}^{\neq} = 2$	5.65 (2.41)	6.95 (1.36)	7.95 (2.63)	6.85 (2.36)
	$\emptyset N_{E+}^{\neq}$	5.52 (2.53)	6.42 (2.10)	7.82 (1.83)	
C = 0 (nichts)	$N_{E+}^{\neq} = 0$	2.00 (1.45)	2.25 (1.77)	3.25 (2.22)	2.50 (1.89)
	$N_{E+}^{\neq} = 1$	2.25 (1.83)	2.70 (1.17)	3.15 (1.90)	2.70 (1.68)
	$N_{E+}^{\neq} = 2$	2.80 (1.88)	3.25 (1.89)	2.30 (2.23)	2.78 (2.01)
	$\emptyset N_{E+}^{\neq}$	2.35 (1.74)	2.73 (1.67)	2.90 (2.13)	

Anmerkungen. Die Tabelle zeigt die durchschnittlichen Bewertungen auf einer Skala von „0 von 10“ bis „10 von 10“ (Häufigkeitsskala mit 11 Stufen) bezüglich der Wahrscheinlichkeit der Anwesenheit des Zieleffekts (Standardabweichung in Klammern). C = Zustand der gemeinsamen Ursache;  $N_{E+}^{\bar{=}}$  = Anzahl der als anwesend beobachteten Effekte der gleichen Kategorie wie der Zieleffekt;  $N_{E+}^{\neq}$  = Anzahl der als anwesend beobachteten Effekte der anderen Kategorie.

Die farblich unterlegten Zellen stellen die durchschnittlichen Bewertungen aggregiert über den jeweils anderen Faktor ( $N_{E+}^{\bar{=}}$  bzw.  $N_{E+}^{\neq}$ ), also marginalisierte Haupteffekte, dar. Die gelb hinterlegten Zellen (Zeilen mit  $\emptyset N_{E+}^{\neq}$ ) entsprechen den als gelbe, gestrichelte Linien dargestellten Datenpunkten in Abbildung 40 (Faktor  $N_{E+}^{\bar{=}}$ ), die grün hinterlegten Zellen (Spalte mit  $\emptyset N_{E+}^{\bar{=}}$ ) entsprechen den als grüne, durchgehende Linien dargestellten Datenpunkten in Abbildung 40 (Faktor  $N_{E+}^{\neq}$ ).

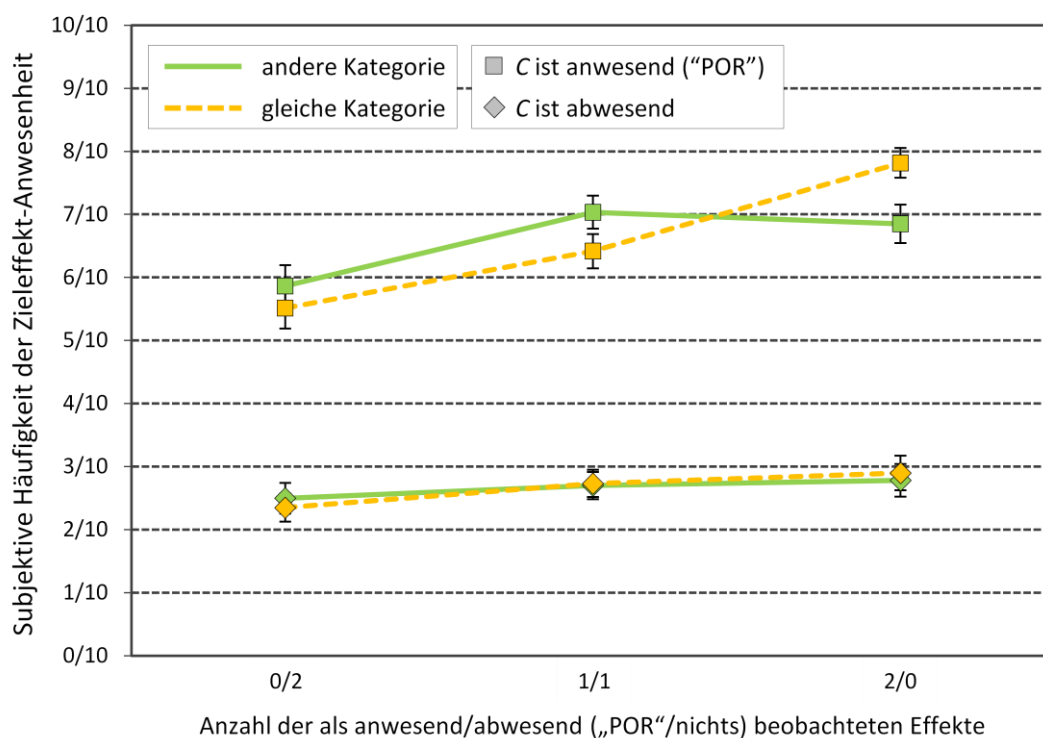


Abbildung 40. Durchschnittliche Ratings bezüglich der Häufigkeit der Anwesenheit des Zieleffekts in zehn fiktiven Situationen gegeben des Status der gemeinsamen Ursache C, der Anzahl der als anwesend beobachteten weiteren Effekte der gleichen Kategorie (Faktor  $N_{E+}^-$ ) sowie der Anzahl der als anwesend beobachteten Effekte der anderen Kategorie (Faktor  $N_{E+}^{\neq}$ ) in Experiment 8 ( $N = 60$ ; balanciertes partiell-konfundiertes randomisiertes Block-Design). Die Fehlerbalken stellen den Standardfehler des Mittelwerts dar.<sup>59</sup>

Die durch die Linien verbundenen Punkte in der Abbildung stellen über die Stufen des jeweils anderen Faktors aggregierte Bewertungen (also Haupteffekte) dar. Die gestrichelte, gelbe Linie mittelt über die Stufen des Faktors  $N_{E+}^{\neq}$ , die durchgezogene, grüne Linie mittelt über die Stufen des Faktors  $N_{E+}^-$ . Man bemerke, dass die grünen und gelben Datenpunkte damit die gleichen Daten repräsentieren, nur auf zwei verschiedenen Arten aggregiert. Dass die beiden Linien sich im Fall der Anwesenheit der Ursache schneiden, kann damit nicht als potentielle Interaktion der beiden Faktoren ( $N_{E+}^- \times N_{E+}^{\neq}$ ) interpretiert werden (siehe für die nicht aggregierten Daten Abbildung M-1 und Abbildung M-2 in Anhang M).

#### 6.4 Diskussion

Die Vorhersagen des Modells konnten in Experiment 8 gut bestätigt werden. Damit wurde demonstriert, dass die Bewertungen der Versuchspersonen hinsichtlich der Anwesenheit eines Zieleffekts in einer *Common-Cause*-Struktur nicht

<sup>59</sup> Der Standardfehler ist nur der Vollständigkeit halber abgetragen. Eine strikte Interpretation im klassischen Sinne ist aufgrund des verwendeten Designs nicht möglich.



nur von den Annahmen über die zugrunde liegenden Kausalprozesse abhängig sind – die wurden im Experiment konstant gehalten –, sondern auch in systematischer Weise von den weiteren Merkmalen der involvierten Effektvariablen: Je ähnlicher ein beobachteter Effekt dem Zieleffekt hinsichtlich weiterer Merkmalen ist, desto größer ist sein Einfluss auf die Beurteilung der Anwesenheit des unbeobachteten Zieleffekts.

Offen gelassen wurde allerdings die nicht uninteressante Frage, wie sich die kausale Relevanz der involvierten Merkmale bestimmt oder wie diese gelernt werden kann. In der Modellierung und der Modellvorhersage wurde von einem *Feature* ausgegangen, dessen kausale Relevanz nahe liegt, über die aber dennoch Unsicherheit besteht. Im Experiment wurde dafür als Eigenschaft die Farbe der Aliens eingesetzt, deren potentielle kausale Relevanz dadurch unterstrichen wurde, dass es sich bei unterschiedlich gefärbten Aliens um unterschiedliche Arten handelt.

Wie würde sich die Vorhersage des Modells verhalten, wenn das Feature nicht kausal relevant wäre? Die Inferenz innerhalb des Modells erfolgt für jede mögliche Partitionierung unter der Annahme einer Abhängigkeit der Effekte innerhalb der zur Partitionierung gehörenden Cluster aber nicht zwischen diesen. Die Relevanz eines Features bestimmt sich daher über die Homogenität der Cluster einer Partitionierung hinsichtlich des Features. Seien zum Beispiel die Likelihoods der möglichen Partitionierungen gegeben des Merkmals (siehe Abbildung 37) alle identisch, also unabhängig vom Merkmal. Da die *A-priori*-Wahrscheinlichkeit einer Partitionierung *per definitionem* auch unabhängig vom Merkmal ist, würde die Mittelung über die Partitionierungen zu einer Zielinferenz führen, die identisch ist mit der Inferenz, die nur einen großen Cluster enthält (Abbildung 36a)<sup>60</sup>, da die Abhängigkeit des Zieleffekts von jedem anderem Effekt gleich groß ist. Dies ist mathematisch äquivalent zum Basismodell. Das heißt, hat die *Common-Cause*-Struktur nur Effekte mit Merkmalen, die als nicht kausal relevant eingestuft werden, dann entspricht die Inferenz dem Basismodell, weil sich die Wirkungen der cluster-spezifischen Einflüsse allesamt ausmitteln – es

---

<sup>60</sup> Dass dem so ist, kann in Anlehnung an die Ableitung in Anhang F relativ einfach gezeigt werden. Die Parameter  $w_{PN}$  und  $w_C$  wären dann natürlich entsprechend ihren Gesamtwirkungen zu bestimmen.

verbleibt nur ein allgemeiner präventiver Einfluss, der auf alle Effekte gleichermaßen wirkt. Dies ist auch nicht überraschend. In den Basisexperimenten wurden Abbildungen von Aliens benutzt, die vielerlei Features hatten. Den Versuchspersonen wurde jedoch explizit mitgeteilt, dass die beschriebenen Eigenschaften des kausalen Systems in gleicher Weise für alle Aliens gelten würden, diese also unabhängig von den visuellen Merkmalen der Aliens seien. Das bedeutet wiederum, dass alle Einflüsse, die sich nur auf Subgruppen von Effekten beziehen, jeweils gleichwahrscheinlich sind und sich damit ausmitteln. Damit ist das in diesem Kapitel vorgestellte Modell eine mathematisch gesehen allgemeinere Fassung des Basismodells, welches dieses als Spezialfall unter der Annahme kausal nicht relevanter Features enthält.

Umfasst das betrachtete kausale System bzw. die involvierten Objekte jedoch kausal relevante Merkmale, äußert sich dies im Modell in der Annahme, dass die Cluster innerhalb der Partitionierungen hinsichtlich des Merkmals homogen sind (z.B. Abbildung 37c), die Wahrscheinlichkeit der Anwesenheit eines Features also entweder nahe 0 (das Merkmal ist im Cluster praktisch immer abwesend) oder nahe 1 (das Merkmal ist im Cluster praktisch immer anwesend) liegt. Dadurch werden die zu den Partitionierungen gehörenden kausalen Strukturen (Abbildung 36) unterschiedlich gewichtet, was zu unterschiedlichen Abhängigkeiten führt. Je geringer die angenommene Homogenität der Cluster im Bezug auf das Merkmal, desto geringer die Unterschiede in den Abhängigkeiten der Effekte hinsichtlich unterschiedlicher Features.

Die Homogenität der Cluster bezüglich eines betrachteten Features wird im Speziellen ausgedrückt über die *A-priori*-Verteilung des Parameters  $\theta_{z_j} \sim \text{Beta}(\alpha_\theta, \beta_\theta)$  mit den Hyperparametern  $\alpha_\theta$  und  $\beta_\theta$ . Für die Bestimmung der Modellvorhersage wurde für die Hyperparameter je  $1/25$  eingesetzt, was einer Beta-Verteilung entspricht, die viel Wahrscheinlichkeitsdichte auf Werte nahe 0 und nahe 1 legt. Die Hyperparameter können dabei natürlich auch auf andere Werte gesetzt werden. Setzt man  $\alpha_\theta$  und  $\beta_\theta$  auf 1 (*flat prior*) entstünde die oben beschriebene Unabhängigkeit, das Merkmal wäre kausal nicht mehr relevant. Setzt man  $\alpha_\theta$  und  $\beta_\theta$  auf sehr hohe, ähnlich große Werte (starke Präferenzen für bestimmte Wahrscheinlichkeiten zwischen 0 und 1) bekäme man den umgekehr-

---

ten Effekt: Da nunmehr besonders inhomogene Cluster ein großes Gewicht bekämen, würde die Inferenz über den Zieleffekt nunmehr stärker von den Effekten abhängen, die dem Zieleffekt möglichst unähnlich sind.<sup>61</sup>

Die Hyperparameter hinsichtlich eines bestimmten Merkmals können dabei – entsprechende Lerndaten vorausgesetzt – im Grunde gelernt werden, da die Parameter im bayesianischen Framework gegeben eines Lerninputs geschätzt werden können. Treten zum Beispiel bei der Beobachtung eines kausalen Systems vornehmlich konditionale Abhängigkeiten im Hinblick auf Effekte auf, die für ein betrachtetes Merkmal die gleiche Ausprägung aufweisen, jedoch keine konditionalen Abhängigkeiten zwischen Effekten unterschiedlicher Merkmalsausprägung, dann handelt es sich um ein relevantes Merkmal mit entsprechend kleinen Hyperparametern (etwa denen für die Modellvorhersage verwendeten). Sind die konditionalen Abhängigkeiten unabhängig vom betrachteten Merkmal, dann ist es nicht relevant mit entsprechenden Hyperparametern nahe 1. In diesem Sinne ist es also möglich, die Relevanz von Merkmalen zumindest dem Grunde nach zu lernen. Fraglich bleibt allerdings, ob die Schätzung der entsprechenden Hyperparameter ohne weitere strukturelle Annahmen mit realistischem Lerninput überhaupt möglich und stabil ist.

## 6.5 Zusammenfassung

In Kapitel 4 wurde auf der Basis der in Kapitel 3 dargestellten Experimente ein Modell entwickelt, dass die Sensitivität der Probanden hinsichtlich ihrer Annahmen über die zugrunde liegenden Kausalprozesse in einer *Common-Cause*-Struktur als adaptive Fehlerattribution implementiert. Im vorliegenden Kapitel wurden diese Überlegungen ergänzt im Hinblick auf die Frage, inwieweit die Inferenzen in kausalen Systemen auch von den weiteren, auf den ersten Blick nicht kausalen Merkmalen (*features*) und damit möglichen Kategorisierungen der involvierten Objekte abhängig sind. Unter der Annahme, dass konditionale Abhängigkeiten nur zwischen Effekten bestehen, die einem Cluster zugeordnet sind, und dass solche Cluster im Hinblick auf kausal relevante Merkmale der kausalen

---

<sup>61</sup> Daher ist die gute Übereinstimmung des Modells mit den in Experiment 8 gefundenen Befunden auch nicht als sonderlich überzeugend im Sinne von Roberts und Pashler (2000) einzustufen, da bei entsprechender Parameterwahl hinsichtlich der Abhängigkeit der Zielinferenz von der Ähnlichkeit der Effekte jedwede Vorhersage modelliert werden könnte.

---

Domäne möglichst homogen sind, wurde eine Erweiterung des in Kapitel 4 vorgestellten Modells entwickelt, welches die Zielinferenz über alle möglichen Partitionierungen der Effekte des Systems bestimmt und über die Unsicherheit bezüglich der Clusterzugehörigkeit integriert. Dieses Modell macht die Vorhersage, dass die Bewertung der Wahrscheinlichkeit der Anwesenheit eines unbeobachteten Zieleffekts in einer *Common-Cause*-Struktur umso stärker von anderen beobachteten Effekten abhängen sollte, je ähnlicher der Zieleffekt diesen anderen Effekten hinsichtlich kausal relevanter Merkmale ist. In einem Experiment, in dem die Versuchspersonen in Anlehnung an die Basisexperimente in einer *Common-Cause*-Struktur mit fünf Effekten die Wahrscheinlichkeit einzuschätzen hatten, mit der ein Zielalien in einer „*Sending*“-Situation an „POR“ denkt, konnte die Vorhersage bestätigt werden: Im Fall der Anwesenheit der Ursache hingen die Bewertungen viel stärker von der Anzahl der „POR“-Gedanken zweier Aliens gleicher Farbe (Alien-Art) ab, als von der Anzahl der „POR“-Gedanken zweier Aliens, die eine andere Farbe als das Zielalien hatten und damit zu einer anderen Alien-Art gehörten. Damit konnte deutlich gezeigt werden, dass die Inferenzen der Versuchspersonen nicht nur von deren Annahmen über die zugrunde liegenden Kausalprozesse, sondern auch von den (als kausal relevant instruierten) Merkmalen der involvierten Objekte abhängig sind – und dass diese Abhängigkeit sehr gut mit der in Kapitel 4 eingeführten Idee einer adaptiven Fehlerattribution beschrieben und erklärt werden kann.

## 7 Gesamtdiskussion und Ausblick

Zum Abschluss der Arbeit sollen die Befunde noch einmal in der Gesamtschau zusammengefasst und diskutiert werden. Darauf aufbauend erfolgt ein kurzer Ausblick möglicher Richtungen zukünftiger Forschung.

### 7.1 Diskussion der Befunde

Ausgangspunkt der Arbeit waren Studien, die die Markov-Bedingung als Default-Annahme im menschlichen Kausaldenken in Frage stellten: Die Einschätzung der Anwesenheit eines nicht beobachteten Zieleffekts in einer *Common-Cause*-Struktur gegeben der An- oder Abwesenheit seiner Ursache hängt im Widerspruch zur Markov-Bedingung sehr deutlich vom Status der weiteren Effekte der gemeinsamen Ursache ab. Die bis dahin vorgebrachten Erklärungen, nach denen die ansonsten Markov-konforme Kausalinferenz mit anderen Inferenzprozessen interagiert (z.B. bezüglich der Typikalität von Exemplaren einer Kategorie oder der Unsicherheit bezüglich der wahren Ausprägung der involvierten Variablen; siehe Rehder & Burnett, 2005 und auch Abschnitt 2.3), wurden als zu „akausal“ und damit nicht weitreichend genug zurückgewiesen. Sowohl im *Underlying-Mechanism*-Modell, welches Markov-Verletzungen durch einen hinzugedachten kategorialen Mechanismus induziert, als auch im *Feature-Uncertainty*-Modell, welches Markov-Verletzungen als Ausdruck der Unsicherheit bezüglich der wahren Ausprägung der Variablen erklärt, ist die eigentliche Kausalinferenz von der Ursache zum Effekt weiterhin Markov-konform. In Abgrenzung dazu wurde ein Modell entwickelt, das direkt an der Kausalinferenz ansetzt und die Annahmen der Probanden hinsichtlich der zugrunde liegenden Kausalprozesse und die damit verbundenen Annahmen über die Fehlerstruktur des Systems berücksichtigt.

In Kapitel 3 wurden drei Experimente vorgestellt, die genau dies gezeigt haben: Die Einschätzungen hinsichtlich der Anwesenheit eines Zieleffekts gegeben des Status der anderen Variablen in einer *Common-Cause*-Struktur werden systematisch von den Annahmen der Versuchspersonen über die zugrunde liegenden Kausalprozesse beeinflusst. Hierzu wurden die Probanden in ein *Common-Cause*-Szenario auf der Basis der „*Mind-Reader-Alien*“-Aufgabe von Steyvers et al. (2003) und der Aufgabe von Rehder und Burnett (2005) eingeführt, das es ermöglichte, diese Annahmen zu manipulieren. Es zeigte sich, dass die Versuchs-

personen bei der Beurteilung der Anwesenheit eines nicht beobachteten Zieleffekts gegeben der Anwesenheit der gemeinsamen Ursache deutlich sensitiver hinsichtlich des Zustands weiterer Effekte dieser gemeinsamen Ursache waren, wenn die gemeinsame Ursache als Gedankensender („*Sending*“-Bedingung) als wenn die Effekte als Gedankenleser („*Reading*“-Bedingung) instruiert wurden. In der „*Sending*“-Bedingung stiegen die Beurteilungen hinsichtlich der Anwesenheit des Zieleffekts sehr stark mit steigender Anzahl weiterer als anwesend beobachteter Effekte an, während die Abhängigkeit in der „*Reading*“-Bedingung deutlich geringer ausgeprägt war. Im Fall der Abwesenheit der Ursache, also wenn diese kausal nicht aktiv war, waren die Bewertungen wie erwartet unabhängig von der Manipulation hinsichtlich der zugrunde liegenden Kausalprozesse.

Diese Befunde zeigen deutlich, dass die Probanden bei ihren Kausalinferenzen sensitiv sind für die dem kausalen System zugrunde liegenden Prozesse und dass Markov-Verletzungen damit in der Kausalinferenz selbst begründet sind. Dies weicht deutlich von den bisherigen Theorien ab, die – wie oben beschrieben – Markov-Verletzungen als Ergebnis einer Interaktion zwischen Markov-konformer Kausalinferenz und anderen Inferenzprozessen erklärt haben, Markov-Verletzungen damit also gerade nicht als „kausales Phänomen“ betrachten.

In Experiment 3 wurde explizit getestet, ob sich die Annahmen bezüglich der Kausalprozesse in unterschiedlichen Annahmen über die zugrunde liegende Fehlerstruktur des kausalen Systems äußern. Im Fall der Anwesenheit der Ursache, jedoch der Abwesenheit aller ihrer beobachteten Effekte attribuierten die Versuchspersonen dieses Problem in der „*Sending*“-Bedingung auf die Ursache, in der „*Reading*“-Bedingung jedoch auf die Effekte. Der Befund scheint damit auf eben diese unterschiedliche Fehlerattribution zurückzuführen zu sein. Als formale Erklärung wurde in Kapitel 4 ein bayesianisches Modell hergeleitet, das die Inferenz als adaptive Fehlerattribution modelliert. Dazu wurde dem Bayes-Netz, welches das kausale System (der beobachtbaren Variablen) repräsentiert, eine zusätzliche Fehlerquelle hinzugefügt, die gleichermaßen präventiv auf alle Links innerhalb einer *Common-Cause*-Struktur wirkt. Unterschiedliche Annahmen über die Stärke dieser gemeinsamen Fehlerquelle führen in diesem Modell zu unter-

schiedlich starken Abhängigkeiten der kausalen Relationen und damit zu unterschiedlich starken Markov-Verletzungen auf der Ebene der beobachteten Variablen. Die modellierte Zielinferenz kann dabei als zweistufiger Prozess gedacht werden: In einem ersten Schritt wird auf die Anwesenheit oder die Abwesenheit der gemeinsamen Fehlerursache geschlossen, je nachdem wie viele beobachtete Effekte der anwesenden Ursache abwesend sind – je mehr, desto unwahrscheinlicher, dass dies rein zufällig durch die unabhängigen Fehlerquellen zustande gekommen ist. In einem zweiten Schritt wird dann auf der Basis dieser Annahmen über die Anwesenheit der gemeinsamen Fehlerquelle auf die Wahrscheinlichkeit der Anwesenheit des Zieleffekts geschlossen.

Aus der Idee der Einführung einer gemeinsamen Fehlerquelle, die auf alle von der Ursache ausgehenden kausalen Links in der betrachteten Struktur wirkt, ergeben sich weitere Vorhersagen, die in weiteren Experimenten getestet wurden. Wendet man das Modell auf eine *Common-Cause*-Struktur an, in der beide Zustände der gemeinsamen Ursache kausal aktiv sind (siehe Abschnitt 5.1), sagt das Modell für die Bewertung der Anwesenheit eines unbeobachteten Zieleffekts eine Abhängigkeit vom Status der weiteren Effekte für *beide* Zustände der Ursache voraus und, dass diese Abhängigkeit wie im Basismodell in einem kausalen System mit starker gemeinsamer Fehlerursache deutlich größer ausfallen sollte, als in einem System mit schwacher gemeinsamer Fehlerursache. Diese Vorhersagen wurden experimentell in einem für den Fall zweier aktiver Zustände der Ursache angepassten, auf dem Paradigma der Basisexperimente beruhenden Experiment getestet. Hierfür wurden beide Gedanken der Außerirdischen („TUS“ und „POR“ statt nichts und „POR“) als kausal aktiv beschrieben und mit der bekannten „*Reading*“-vs.-„*Sending*“-Manipulation – also einer eher effektseitigen oder einer eher ursacheseitigen Fehlerattribution – kombiniert. Wie vorhergesagt, zeigte sich eine Abhängigkeit vom Status der weiteren Effekte für beide Zustände der Ursache, und zwar deutlich stärker in der „*Sending*“-Bedingung. Dieses Ergebnis ist dabei besonders interessant, als dass die in den Basisexperimenten gefundene Asymmetrie zwischen Anwesenheit und Abwesenheit der Ursache damit wirklich auf den unterschiedlichen kausalen Status und nicht zum Beispiel auf die unterschiedlich instruierten Basisraten der beiden Zustände zurückzuführen

---

ren ist. Da die Erklärungsversuche von Rehder und Burnett (2005) nicht an der Kausalinferenz ansetzen und damit auch nicht sensitiv sein können für den kausalen Status der verschiedenen Ausprägungen der Ursache, versagen diese, wenn sie diesen gefundenen Unterschied erklären sollen.

Auch auf andere Kausalstrukturen lässt sich die Idee einer gemeinsamen, an die Ursache gebundenen Fehlerquelle anwenden (siehe Abschnitt 5.2). In einer kausalen Kette, wie sie auch bereits von Rehder und Burnett (2005) untersucht wurde, sollte eine Manipulation hinsichtlich der zugrunde liegenden kausalen Prozesse, sofern diese unterschiedliche Fehlerattributionen nahe legen, keinen Einfluss haben. Bei der Bewertung der Anwesenheit des ultimativen Effekts in einer kausalen Kette kommt es nicht darauf an, ob mögliche Fehler ursache- oder effektseitig attribuiert werden, maßgeblich ist nur die Gesamtstärke der Ursache auf den Zieleffekt. Die Vorhersagen wurden experimentell dergestalt getestet, dass das in den Basisexperimenten verwendete Paradigma für eine kausale Kette angepasst wurde, indem den Probanden (entsprechend je nach Bedingung) erklärt wurde, dass in einer Reihe von Außerirdischen jeweils links stehende Alien verursache die Gedanken des jeweils rechts stehenden. Wie vorhergesagt unterschieden sich die Bewertungen der Probanden hinsichtlich der Anwesenheit des ultimativen Effekts *nicht* im Hinblick auf die „*Reading*“-vs.-„*Sending*“-Manipulation und damit gerade nicht bezüglich einer eher ursache- oder effektseitigen Fehlerattribution.

Da das in dieser Arbeit vorgestellte Modell den zugrunde liegenden Inferenzprozess, der die konditionalen Abhängigkeiten erzeugt, als adaptive Fehlerattribution beschreibt, wurde in Abschnitt 5.3 des Weiteren getestet, welchen Einfluss es hat, wenn in eine *Common-Cause*-Struktur weitere beobachtbare präventive Ursachen eingeführt werden, die mit der modellendogenen Fehlerquelle bei eben dieser Fehlerattribution konkurrieren. Wenn in das kausale System eine solche als anwesend beobachtete präventive Ursache eingeführt wird, die nicht auf den Zieleffekt, aber auf alle anderen beobachteten Effekte wirkt, sagt das Modell vorher, dass die Einschätzung bezüglich der Anwesenheit eines unbeobachteten Zieleffekts deutlich weniger von der Anzahl der weiteren als anwesend beobachteten Effekte abhängen sollte, da deren Abwesenheit



vielmehr auf die beobachtete präventive Ursache attribuiert wird. Dies reduziert die Attribution auf die modellendogen angenommene unbeobachtete Fehlerquelle und damit den Einfluss des Zustands der weiteren Effekte auf den Zieeffekt. Diese Vorhersagen wurden in zwei Experimenten getestet, die an die Basisexperimente angelehnt waren und sich auf den „*Sending*“-Fall als zugrunde liegenden Mechanismus mit eher ursacheseitiger Fehlerattribution beschränken (genauerer siehe Abschnitt 5.3): Dabei zeigte sich, dass im Fall der Anwesenheit der gemeinsamen Ursache sowie der Anwesenheit einer präventiven Ursache – in diesem Fall die Einführung zweier Wände, die die beiden nicht Zieeffekt-Aliens vom Gedanken sendenden Alien abschirmten – die Einschätzung der Anwesenheit des unbeobachteten Zieeffekts von der Anzahl der als anwesend beobachteten weiteren Effekte deutlich geringer abhing, als ohne diese präventive Ursache. Damit konnte das Modell auch in einer Situation bestätigt werden, in der die postulierte Fehlerattribution direkt durch die Einführung weiterer präventiver Ursachen manipuliert wurde.

In der Gesamtschau hat sich demnach die in Abschnitt 2.4 entwickelte Idee sehr gut bewährt. Markov-Verletzungen sind demnach nicht Ausdruck einer an sich Markov-konformen Kausalinferenz, die mit anderen Inferenzprozessen interagiert, wie dies noch von Rehder und Burnett (2005) angenommen wurde. Konditionale Abhängigkeiten sind demnach vor allem Ausdruck der Annahmen über die zugrunde liegenden Kausalprozesse und die damit verbundene Fehlerstruktur (für diesbezügliche Einschränkungen siehe weiter unten). Die Kausale-Bayes-Netz-Theorie, wie sie in Abschnitt 2.2 als psychologische Theorie kausalen Denkens eingeführt wurde (siehe u.a. Gopnik et al., 2004; Sloman, 2005), muss daher erweitert werden. Die Markov-Bedingung scheint im Hinblick auf die beobachtbaren Variablen offenkundig nicht Bestandteil der Repräsentation zu sein, auf denen das kognitive System operiert. Das hier vorgeschlagene Modell, welches das Wissen um die zugrunde liegenden kausalen Mechanismen in der Form von Annahmen über die Fehlerstruktur integriert, schlägt damit konzeptionell auch eine erste Brücke zwischen den mechanismusbasierten Ansätzen (siehe Abschnitt 2.1.3), die die Bedeutungen kausaler Mechanismen für das menschliche Kausaldenken hervorheben, und den mit den kovariationsbasierten Ansätzen

verwandten Bayes-Netz-Theorien, die Kausaldenken eher formal auf statistischer Ebene beschreiben.

In diesem Zusammenhang soll allerdings nicht unerwähnt bleiben, dass in der Zwischenzeit in der Arbeitsgruppe um Steve Sloman ähnliche Ideen entwickelt und getestet wurden, die jedoch nicht so weitreichend wie die hier vorgestellten sind. In ihren Untersuchungen haben Walsh und Sloman (2007) ihre Probanden bezüglich einer *Common-Cause*-Struktur mit einer gemeinsamen Ursache  $C$  und zwei Effekten,  $E_1$  und  $E_2$ , dieser Ursache (siehe Abbildung 1b auf Seite 11 entsprechend) in verschiedenen kausalen Szenarien instruiert, z.B. dass das Spielen lauter Musik [ $C$ ] dazu führt, dass sich die linken Nachbarn beschweren [ $E_1$ ] und dass sich die rechten Nachbarn beschweren [ $E_2$ ]. In der Testphase sollten die Versuchspersonen die Wahrscheinlichkeit des Auftretens von Effekt  $E_2$  gegeben der Ursache  $C$  einschätzen, sowohl bevor die Probanden  $E_1$  beobachten konnten, als auch nachdem die Probanden wussten, dass  $E_1$  nicht aufgetreten ist. Dabei zeigte sich, dass die Urteile über den Effekt  $E_2$  nicht unabhängig vom Wissen der Probanden über die Abwesenheit des Effekts  $E_1$  ist. Gegeben des Wissens um die Abwesenheit von  $E_1$  war die Einschätzung der Wahrscheinlichkeit des Auftretens von  $E_2$  deutlich geringer, als vor der Information um die Abwesenheit von  $E_1$  (was der Markov-Bedingung widerspricht). Walsh und Sloman (2007) konnten zeigen, dass die Stärke dieser konditionalen Abhängigkeit davon beeinflusst ist, womit die Probanden es sich erklären, dass der beobachtete Effekt  $E_1$  nicht aufgetreten ist. Als mögliche Erklärung entwickelten Walsh und Sloman die Idee, dass die Probanden *Ad-hoc*-Änderungen an ihrem gerade verwendeten Kausalmodell vornehmen, wenn sie Widersprüche entdecken, und dass sie in der speziellen Situation mit diesem geänderten Modell weiter operieren. Wird z.B. eine Erklärung generiert, die sich auf eine verhindernde Bedingung (*disabling condition*) bezieht, die beide Effekte umfasst und damit beide Effekte von der Ursache entkoppelt, sollte sich dies in einer starken konditionalen Abhängigkeit niederschlagen. Eine Erklärung für die Abwesenheit von  $E_1$ , die eine verhindernde Bedingung enthält, die nur  $E_1$  betrifft, sollte entsprechend nicht die Anwesenheit von  $E_2$  betreffen und sich damit auch nicht in einer konditionalen Abhängigkeit äußern. Um dies zu testen, haben Walsh und Sloman (2007) in einem weiteren Experi-

ment die Auswirkungen verschiedener Erklärungsmöglichkeiten für das Nicht-Auftreten des beobachteten Effekts explizit manipuliert. Dazu nutzten sie zum einen als Typ von Erklärungen solche, die sich – wie eben beschrieben – nur auf die Verhinderung des betroffenen Effekts beziehen, wie z.B. dass die linken Nachbarn sich nicht über die laute Musik beschwerten, weil sie ausgegangen sind. Zum anderen nutzten sie Erklärungen, die in ihrer verhindernden Wirkung beide Effekte betreffen, z.B. dass sich die linken Nachbarn nicht über die laute Musik beschwerten, weil man alle seine Nachbarn zur Party eingeladen hat (auf der die laute Musik gespielt wird). Über eine Reihe von Szenarien wurden die Versuchspersonen dann wieder gebeten, die Wahrscheinlichkeit der Anwesenheit des unbeobachteten Effekts einzuschätzen und zwar sowohl bevor ihnen die Abwesenheit des zweiten Effekts mitgeteilt wurde, als auch danach – nur dass ihnen diesmal auch explizit eine der Erklärungen für die Abwesenheit vorgelegt wurde. Wie erwartet, führten die Erklärungen zweiten Typs (beide Effekte betroffen) zu deutlich stärkeren konditionalen Abhängigkeiten, als die Erklärungen ersten Typs (nur der beobachtete Effekt betroffen). Nichtsdestotrotz waren die Bewertungen der Versuchspersonen hinsichtlich der Anwesenheit des unbeobachteten Effekts auch im letzteren Fall deutlich vom Wissen um den Status des anderen Effekts abhängig. Walsh und Sloman schlossen daraus, dass menschliche Kausalurteile grundsätzlich nicht die Markov-Bedingung erfüllen und dass kausale Bayes-Netze daher möglicherweise keine geeignete theoretische Basis darstellen, um menschliches Kausaldenken zu beschreiben.

Die Idee von Walsh und Sloman (2007) zielt damit in die gleiche Richtung, wie das in dieser Arbeit entwickelte Modell: Die Inferenzen sind durch Annahmen über mögliche Fehler, hier im Besonderen verhindernde Bedingungen, beeinflusst. Die rein strukturell-qualitative Erklärung, die Markov-Verletzungen als Folge von *Ad-hoc*-Änderungen an dem der Inferenz zugrunde liegenden Kausalmodell im Hinblick auf das Einfügen dieser verhindernden Bedingungen beschreibt, greift jedoch zu kurz. Es kann vor allem die von Walsh und Sloman gefundenen Verletzungen nicht erklären, die auch dann noch auftreten, wenn gar keine verhindernde Bedingung, die den Zieleffekt mit umfasst, instruiert bzw. von den Probanden vorgebracht wurde. Damit geht das hier vorgestellte Modell

---

deutlich weiter, da es die Inferenzprozesse quantitativ als adaptive Fehlerattribution beschreibt, die nicht abrupt (also über distinkte Strukturen), sondern fließend innerhalb ein und derselben Kausalstruktur erfolgt und daher auch diese Befunde erklären kann.

Es soll allerdings nicht verschwiegen werden, dass auch das in dieser Arbeit vorgestellte Modell nicht ausreichend scheint. Mit den oben beschriebenen Experimenten konnte zwar deutlich nachgewiesen werden, dass Markov-Verletzungen vornehmlich durch die auf den Annahmen hinsichtlich der zugrunde liegenden Kausalprozesse implizierten Fehlerstruktur zurückzuführen und damit in Abgrenzung zu den Erklärungsversuchen von Rehder und Burnett (2005) ein „kausales Phänomen“ sind. Dies reicht aber offenkundig als alleinige Erklärung nicht aus. In den beiden Basisexperimenten Experiment 2 und Experiment 3 konnte eine leichte konditionale Abhängigkeit auch für den Fall der Abwesenheit der Ursache nachgewiesen werden, auch in Experiment 1 ist sie der Tendenz nach vorhanden. Dies widerspricht dem Modell, da die Ursache im Falle ihrer Abwesenheit kausal nicht aktiv ist und damit auch keine kausalen Prozesse (innerhalb des betrachteten Systems) eine Rolle spielen können. Das gleiche Problem trat in Experiment 5 – also der Anwendung des Modells auf eine kausale Kette – auf. Zwar gab es wie vorhergesagt keinen Einfluss der Manipulation hinsichtlich der Annahmen über die Prozesse, dennoch traten konditionale Abhängigkeiten sowohl im Falle der Anwesenheit als auch im Falle der Abwesenheit der Ursache auf, die vom Modell nicht erklärt werden können. Dies spricht dafür, dass die Probanden auch weitere Faktoren berücksichtigen, die nichts mit dem kausalen System an sich zu tun haben. Bezeichnenderweise gab es diese Abhängigkeiten im Falle der Abwesenheit der Ursache *nicht* in Experiment 6, in welchem das kausale System (hier: die vier Aliens) von einer Wand umschlossen und damit von der Umwelt quasi abgeschirmt waren. Conclusio kann damit also nur sein, dass die Annahmen hinsichtlich der zugrunde liegenden Kausalprozesse und die dadurch implizierte Fehlerstruktur bei Kausalinferenzen, wie sie hier untersucht worden sind, eine sehr wichtige, aber je nach Kontext nicht die einzige Rolle spielen. Andere Prozesse, wie sie z.B. von Rehder und Burnett (2005) vorgeschlagen worden sind, sollen also nicht ausgeschlossen werden.

Ganz außen vorgelassen worden, aber nicht minder interessant, ist in der vorliegenden Arbeit dabei die Frage, *wie* man eigentlich aus seinem möglichen Wissen um die involvierten Kausalprozesse die für die Inferenz relevanten Parameter der Fehlerstruktur ableiten kann. Für die Generierung der Modellvorhersagen ist hierfür die entsprechende Stärke der präventiven Ursache je nach Bedingung auf verschiedene Werte gesetzt worden. Die Ableitung dieser Stärke aus den Kausalprozessen ist derzeit also *a priori* und nicht Bestandteil des Modells. Das lässt sich zwar leicht dadurch rechtfertigen, dass aus computationaler und damit Modellierungssicht als erster Schritt die statistischen Inferenzen abgebildet werden müssen, die dem Schlussprozess zugrunde liegen, nur als nächster Schritt scheint es notwendig, diesen mehr qualitativen Teil des Inferenzprozesses in das Modell zu integrieren. Möglichkeiten hierfür sollen im kommenden Abschnitt skizziert werden.

## 7.2 Ausblick

Bereits in Kapitel 4 ist ausgeführt worden, dass Bayes-Netze nur statistische Abhängigkeiten auf der Ebene von Ereignissen repräsentieren. In Kapitel 6, in dem das Modell auf Ereignisse mit unterschiedlichen Merkmalen erweitert worden war, traten neben diese rein statistischen Informationen auch weitere Informationen, die kausal gesehen nicht direkt relevant sind, die sich aber indirekt – z.B. über mögliche Kategorisierungen der Ereignisse – auf die statistischen Abhängigkeiten auswirken. Dies wurde im Rahmen eines hierarchischen Bayes-Ansatzes unter der Annahme modelliert, dass die durch die ursacheseitig eingeführte Fehlerquelle implizierten konditionalen Abhängigkeiten nur zwischen Effekten bestehen, die einem Cluster zugeordnet sind, und dass solche Cluster im Hinblick auf kausal relevante Merkmale der involvierten Objekte möglichst homogen sind. Die letztendliche Inferenz fand dann über alle möglichen Partitionierungen der Effekte statt, indem diese ausintegriert wurden. Ähnliche, auf der Basis von Ereignismerkmalen operierende hierarchische Bayes-Ansätze finden sich auch bereits bei Kemp, Goodman et al. (2007) sowie Kemp, Shafto, Berke und Tenenbaum (2007).

Das eigentliche Problem hierbei ist, dass im Prinzip nicht die Ereignisse die beschriebenen Merkmale aufweisen, sondern die Objekte, die mit den Ereignis-

sen in Verbindung stehen. In den meisten Fällen ist diese Unterscheidung nicht wichtig, oft ist das statistische relevante Ereignis schlicht die Anwesenheit vs. Abwesenheit des Objekts (z.B. bei einem Virus). In den vorgestellten Experimenten ist dies schon etwas komplexer: Die Ereignisse sind die gedachten Wörter, zwischen denen statistische Abhängigkeiten bestehen, die Objekte sind die Außerirdischen, die – wie in Experiment 8 – auch Merkmale aufweisen können. Die weiteren angenommenen, unbeobachteten Ereignisse, wie z.B. die Unaufmerksamkeit des Senders oder die Unaufmerksamkeit der Leser sind ebenfalls mit diesen Objekten assoziiert und die Stärke dieser (präventiven) Ereignisse leitet sich offenkundig aus der Intuition über den aktiven Teil des Mechanismus ab. Im „*Sending*“-Fall ist das Ursache-Objekt aktiv (und kann daher versagen), im „*Reading*“-Fall sind die Effekt-Objekte aktiv (und können daher versagen).

Um dieses Hintergrundwissen computational abbilden zu können, scheint es nur folgerichtig, explizit zwischen der rein statistischen Ebene der Ereignisse (und deren statistischen Abhängigkeiten) und der kausalen Hintergrundebene zu unterscheiden und diese im Wege eines hierarchischen Bayes-Modells miteinander zu verbinden. Die kausale Hintergrundebene umfasst dabei das Wissen um Objekte mit Merkmalen und deren Interaktionen mit anderen Objekten. Die Differenzierung erlaubt es dabei, auch zusätzliche Entitäten in diese Ebene aufzunehmen und dort zu beschreiben, so zum Beispiel (und vor allem) das abstrakte oder auch konkrete Wissen um Mechanismen und deren Merkmale sowie relevante Kontextfaktoren (wie zum Beispiel präventive Ursachen). Kausalwissen wäre demnach auf dieser Hintergrundebene repräsentiert und Kausallernen würde auch auf dieser Ebene stattfinden (siehe Ahn & Kalish, 2000; ohne dass diese jedoch eine formale Lösung anbieten). Inferenzen über Ereignisse lassen sich dann computational als Inferenz in einem Bayes-Netz darstellen, dass aus dieser Hintergrundebene eben für diese Inferenz abgeleitet wurde, genau wie das Beobachten von statistischen Ereignissen und deren Abhängigkeiten sich wieder zurück auswirken würde auf die Hintergrundebene.<sup>62</sup> Dies ließe sich in einer Grammatik formalisieren, ähnlich wie sie bereits von Griffiths und Tenenbaum (2007b) und Tenenbaum et al. (2007) allerdings auf der rein statisti-

---

<sup>62</sup> Zum Beispiel im Hinblick darauf, ob beobachtete Abhängigkeiten nur Koinzidenzen darstellen oder kausal begründet sind.

schen Ebene genutzt wurde (ohne die eben vorgeschlagene Unterscheidung vorzunehmen). Mit der Unterscheidung dieser beiden Ebenen und deren Verknüpfung wäre dann auch auf formaler Ebene die Verbindung zwischen den kovariationsbasierten Ansätzen/Bayes-Netzen und den mechanismusbasierten Ansätzen gelungen, die beide für sich genommen eine bedeutende Rolle spielen, deren Vereinigung aber bis dato noch nicht wirklich gelungen ist (siehe z.B. Newsome, 2003, aber auch Danks, 2005).

Zusammenfassend kann damit bemerkt werden, dass der aktuelle Forschungsstand im Bereich des menschlichen Kausaldenkens bei Weitem noch nicht die Komplexität des Untersuchungsgegenstands widerspiegelt. Wie der Überblick in Kapitel 2 jedoch zeigt, sind in den letzten Jahren enorme Fortschritte zu verzeichnen gewesen. Auch in dieser Arbeit wurde versucht, einen kleinen und hoffentlich nicht unbedeutenden Schritt beizutragen, einer adäquateren Beschreibung und Erklärung dieses komplexen psychologischen Phänomens wie auch computationalen Problems näher zu kommen.

## 8 Zusammenfassung

Die Fähigkeit, kausale Beziehungen in der Welt zu entdecken und das Wissen um diese nutzbar zu machen, ist eine zentrale Kompetenz, um in der Umwelt erfolgreich agieren zu können. Eine bedeutende Rolle in der aktuellen psychologischen Forschung um eben dieses Kausalwissen spielt die Theorie der kausalen Bayes-Netze, die Ursache-Wirkungs-Beziehungen in gerichteten Graphen formalisiert. Zentrale Annahme dieses Ansatzes ist die *Markov-Bedingung*, nach der eine Variable konditionalisiert auf ihre direkten Ursachen unabhängig von allen anderen, nicht nachfolgenden Variablen des Systems ist. Neuere Forschung nährt jedoch Zweifel an der Markov-Bedingung als Teil einer psychologischen Theorie des kausalen Denkens (siehe v.a. Rehder & Burnett, 2005). So hängt die Beurteilung der Anwesenheit eines Effektes in einer *Common-Cause*-Struktur gegeben der An- oder Abwesenheit seiner Ursache im Widerspruch zur Markov-Bedingung sehr wohl und gleichsam auch sehr deutlich vom Status der weiteren Effekte der gemeinsamen Ursache ab.

In der vorliegenden Arbeit wird empirisch gezeigt, dass die Probanden bei solchen Kausalinferenzen systematisch weiteres Wissen, insbesondere um die zugrunde liegenden kausalen Prozesse, miteinbeziehen und dies zu „Markov-Verletzungen“ führt. Aufbauend darauf wird ein computationales Modell entwickelt, welches klassische Bayes-Netze um eine ursacheseitige Fehlerkomponente erweitert und dabei den Schlussprozess als adaptive Fehlerattribution implementiert. Dieses Basismodell wird dann in drei Anwendungsbereichen empirisch geprüft. Des Weiteren wird gezeigt, dass die Inferenzen auch von den Merkmalen der involvierten Objekte und damit von möglichen Kategorisierungen der Effektivvariablen abhängig sind, und eine entsprechende Modellerweiterung entwickelt, die die Zielinferenz über alle möglichen Partitionierungen der Effekte bestimmt und dabei über die Unsicherheit bezüglich der Clusterzugehörigkeit integriert. Abschließend werden mögliche Konsequenzen der Befunde diskutiert und ein umfassenderes, computationales Modell des Kausaldenkens skizziert, welches zwischen einer statistischen Ebene der Ereignisse und einer kausalen Hintergrundebene, in der die involvierten Objekte und Mechanismen repräsentiert sind, unterscheidet.



## 9 Literaturverzeichnis

- Ahn, W.-K. (1998). Why are different features central for natural kinds and artifacts? The role of causal status in determining feature centrality. *Cognition*, *69*, 135–178.
- Ahn, W.-K. & Bailenson, J. (1996). Causal attribution as a search for underlying mechanisms: An explanation of the conjunction fallacy and the discounting principle. *Cognitive Psychology*, *31*, 83–123.
- Ahn, W.-K. & Kalish, C. W. (2000). The role of mechanism beliefs in causal reasoning. In F. C. Keil & R. A. Wilson (Eds.), *Explanation and Cognition*. Cambridge, MA: MIT Press.
- Ahn, W.-K., Kalish, C. W., Gelman, S., Medin, D. L., Luhmann, C., Atran, S. et al. (2001). Why essences are essential in the psychology of concepts. *Cognition*, *82*, 59–69.
- Ahn, W.-K., Kalish, C. W., Medin, D. L. & Gelman, S. A. (1995). The role of covariation versus mechanism information in causal attribution. *Cognition*, *54*, 299–352.
- Allan, L. G. & Jenkins, H. M. (1980). The judgment of contingency and the nature of the response alternatives. *Canadian Journal of Psychology*, *34*, 1–11.
- Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York: Springer.
- Blaisdell, A. P., Sawa, K., Leising, K. J. & Waldmann, M. R. (2006). Causal reasoning in rats. *Science*, *311*, 1020–1022.
- Brase, G. L. (2002). Which statistical formats facilitate what decisions? The perception and influence of different statistical information formats. *Journal of Behavioral Decision Making*, *15*, 381–401.
- Brighton, H. & Gigerenzer, G. (2008). Bayesian brains and cognitive mechanism: Harmony or dissonance? In N. Chater & M. Oaksford (Eds.), *The probabilistic mind: Prospects for Bayesian cognitive science* (S. 189–208). New York: Oxford University Press.
- Buehner, M. J. (2005). Contiguity and covariation in human causal inference. *Learning & Behavior*, *33* (2), 230–238.

- 
- Buehner, M. J., Cheng, P. W. & Clifford, D. (2003). From covariation to causation: A test of the assumption of causal power. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29 (6), 1119–1140.
- Bullock, M. & Gelman, R. (1979). Preschool children's assumptions about cause and effect: Temporal ordering. *Child Development*, 50, 89–96.
- Campbell, J. (2007). An interventionist approach to causation in psychology. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (S. 58–66). Oxford: Oxford University Press.
- Cartwright, N. (1983). *How the laws of physics lie*. Oxford: Clarendon Press.
- Cartwright, N. (1989). *Nature's capacities and their measurement*. Oxford: Clarendon Press.
- Cartwright, N. (2001). What is wrong with Bayes nets? *The Monist*, 84 (2), 242–264.
- Cartwright, N. (2002). Against modularity, the causal Markov condition, and any link between the two: Comments on Hausman and Woodward. *The British Journal for the Philosophy of Science*, 53 (3), 411–453.
- Cartwright, N. (2007). *Hunting causes and using them: Approaches in Philosophy and Economics*. Cambridge, UK: Cambridge University Press.
- Chater, N. & Manning, C. D. (2006). Probabilistic models of language processing and acquisition. *Trends in Cognitive Science*, 10 (7), 335–344.
- Chater, N. & Oaksford, M. (2008a). The probabilistic mind: Prospects for a Bayesian cognitive science. In N. Chater & M. Oaksford (Eds.), *The probabilistic mind: Prospects for Bayesian cognitive science* (S. 3–31). New York: Oxford University Press.
- Chater, N. & Oaksford, M. (2008b). The probabilistic mind: Where next? In N. Chater & M. Oaksford (Eds.), *The probabilistic mind: Prospects for Bayesian cognitive science* (S. 501–514). New York: Oxford University Press.
- Chater, N., Tenenbaum, J. B. & Yuille, A. (2006). Probabilistic models of cognition: Conceptual foundations. *Trends in Cognitive Science*, 10 (7), 287–291.
- Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, 104 (2), 367–405.

- 
- Cheng, P. W. & Novick, L. R. (1990). A probabilistic contrast model of causal induction. *Journal of Personality and Social Psychology*, *58* (4), 545–567.
- Cheng, P. W., Novick, L. R., Liljeholm, M. & Ford, C. (2007). Explaining four psychological asymmetries in causal reasoning: Implications of causal assumption for coherence. In M. O'Rourke (Ed.), *Topics in contemporary philosophy, Vol. 4: Explanation and causation* (S. 1–32). London: MIT Press.
- Collins, D. J. & Shanks, D. R. (2006). Conformity to the power PC theory of causal induction depends on the type of probe question. *The Quarterly Journal of Experimental Psychology*, *59* (2), 225–232.
- Danks, D. (2003). Equilibria of the Rescorla-Wagner model. *Journal of Mathematical Psychology*, *47*, 109–121.
- Danks, D. (2005). The supposed competition between theories of human causal inference. *Philosophical Psychology*, *18* (2), 259–272.
- Danks, D. (2007). Theory unification and graphical models in human categorization. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (S. 173–189). Oxford: Oxford University Press.
- Danks, D. (2008). Rational analyses, instrumentalism, and implementations. In N. Chater & M. Oaksford (Eds.), *The probabilistic mind: Prospects for Bayesian cognitive science* (S. 59–75). New York: Oxford University Press.
- Dickinson, A. (2001). Causal learning: An associative analysis. *Quarterly Journal of Experimental Psychology*, *54B*, 3–25.
- Gelman, A., Carlin, J. B., Stern, H. S. & Rubin, D. B. (2004). *Bayesian data analysis* (2<sup>nd</sup> edition). London: Chapman & Hall.
- Gigerenzer, G. & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, *102*, 684–704.
- Gigerenzer, G. & Hoffrage, U. (1999). Overcoming difficulties in Bayesian reasoning: A reply to Lewis and Keren (1999) and Mellers and McGraw (1999). *Psychological Review*, *106*, 425–430.
- Glymour, C. (1998). Psychological and normative theories of causal power and the probability of causes. In G. Cooper & S. Moral (Eds.), *Uncertainty in Artificial Intelligence* (S. 166–172). San Francisco, CA: Morgan Kaufmann.

- 
- Glymour, C. (2001). *The mind's arrows: Bayes nets and graphical causal models in psychology*. Cambridge, MA: MIT Press.
- Glymour, C. (2003). Learning, prediction and causal Bayes nets. *Trends in Cognitive Science*, 7, 43–48.
- Glymour, C. (2007). Learning the structure of deterministic systems. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (S. 231–240). Oxford: Oxford University Press.
- Goedert, K. M. & Spellman, B. A. (2005). Nonnormative discounting: There is more to cue interaction effects than controlling for alternative causes. *Learning & Behavior*, 33 (2), 197–210.
- Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T. & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review*, 111, 3–32.
- Gopnik, A., Sobel, D. M., Schulz, L. E. & Glymour, C. (2001). Causal learning mechanisms in very young children: Two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology*, 37 (5), 620–629.
- Griffiths, T. L., Kemp, C., and Tenenbaum, J. B. (2008). Bayesian models of cognition. In R. Sun (Ed.), *The Cambridge handbook of computational cognitive modeling*. Cambridge, UK: Cambridge University Press.
- Griffiths, T. L., Steyvers, M. & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review*, 114, 211–244.
- Griffiths, T. L. & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology*, 51, 354–384.
- Griffiths, T. L. & Tenenbaum, J. B. (2007a). From mere coincidences to meaningful discoveries. *Cognition*, 103, 180–226.
- Griffiths, T. L. & Tenenbaum, J. B. (2007b). Two proposals for causal grammars. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (S. 323–345). Oxford: Oxford University Press.
- Griffiths, T. L. & Yuille, A. (2008). A primer on probabilistic inference. In N. Chater & M. Oaksford (Eds.), *The probabilistic mind: Prospects for Bayesian cognitive science* (S. 33–57). New York: Oxford University Press.

- 
- Hagmayer, Y. & Meder, B. (2008). Causal learning through repeated decision making. In B. C. Love, K. McRae & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (S. 179–184). Austin, TX: Cognitive Science Society.
- Hagmayer, Y., Sloman, S. A., Lagnado, D. A. & Waldmann, M. R. (2007). Causal reasoning through intervention. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (S. 86–100). Oxford: Oxford University Press.
- Hagmayer, Y. & Waldmann, M. R. (2002). How temporal assumptions influence causal judgments. *Memory & Cognition*, 30 (7), 1128–1137.
- Hagmayer, Y. & Waldmann, M. R. (2006). Kausales Denken. In J. Funke (Hrsg.), *Enzyklopädie der Psychologie „Denken und Problemlösen“*, Band C/II/8 (S. 87–166). Göttingen: Hogrefe.
- Hausman, D. M. & Woodward, J. (1999). Independence, invariance, and the causal Markov condition. *British Journal for the Philosophy of Science*, 50, 521–583.
- Hausman, D. M. & Woodward, J. (2004). Modularity and the causal Markov condition: A restatement. *British Journal for the Philosophy of Science*, 55, 147–161.
- Heckerman, D. (1999). A tutorial on learning with Bayesian networks. In M. Jordan (Ed.), *Learning in graphical models*. Cambridge, MA: MIT Press.
- Heckerman, D., Meek, C. & Cooper, G. (1999). A Bayesian approach to causal discovery. In C. Glymour & G. Cooper (Eds.), *Computation, causation and discovery*. Cambridge, MA: MIT Press.
- Jenkins, H. M. & Ward, W. C. (1965). Judgment of contingency between responses and outcomes. *Psychological Monographs*, 79.
- Kemp, C., Goodman, N. D. & Tenenbaum, J. B. (2007). Learning causal schemata. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th Annual Cognitive Science Society* (S. 64–70). Austin, TX: Cognitive Science Society.
- Kemp, C., Perfors, A. & Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science*, 10 (3), 307–321.

- 
- Kemp, C., Shafto, P., Berke, A. & Tenenbaum, J. B. (2007). Combining causal and similarity-based reasoning. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th Annual Cognitive Science Society* (S. 64–70). Austin, TX: Cognitive Science Society.
- Kemp, C. & Tenenbaum, J. B. (2008). The discovery of structural form. *Proceedings of the National Academy of Sciences*, *105* (31), 10687–10692.
- Kemp, C. & Tenenbaum, J. B. (in press). Structured statistical models of inductive reasoning. *Psychological Review*.
- Kirk, R. E. (1995). *Experimental Design: Procedures for the behavioral sciences* (3<sup>rd</sup> ed.), Pacific Grove, CA: Brooks/Cole.
- Koslowski, B., Okagaki, L., Lorenz, C. & Umbach, D. (1989). When covariation is not enough: The role of causal mechanism, sampling method and sample size in causal reasoning. *Child Development*, *60* (6), 1316–1328.
- Krynski, T. R. (2006). The role of temporal factors and prior knowledge in causal learning and judgment. Unveröffentlichte Dissertation, Massachusetts Institute of Technology, Cambridge, USA.
- Kushnir, T. & Gopnik, A. (2007). Conditional probability versus spatial contiguity in causal learning: Preschoolers use new contingency to overcome prior spatial assumptions. *Developmental Psychology*, *43* (1), 186–196.
- Lagnado, D. A. & Sloman, S. A. (2002). Learning causal structure. In W. Gray & C. D. Schunn (Eds.), *Proceedings of the 24th Annual Conference of the Cognitive Science Society* (S. 560–565). Mahwah, NJ: Erlbaum.
- Lagnado, D. A. & Sloman, S. A. (2004). The advantage of timely intervention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30* (4), 856–876.
- Lagnado, D. A. & Sloman, S. A. (2006). Time as a guide to cause. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32* (3), 451–460.
- Lagnado, D. A., Waldmann, M. R., Hagmayer, Y. & Sloman, S. A. (2007). Beyond covariation: Cues to causal structure. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (S. 154–172). Oxford: Oxford University Press.

- 
- Leising, K. J., Wong, J., Waldmann, M. R. & Blaisdell, A. P. (2008). The special status of action in causal reasoning in rats. *Journal of Experimental Psychology: General*, *137* (3), 514–527.
- Lewis, C. & Keren, G. (1999). On the difficulties underlying Bayesian reasoning: Comment on Gigerenzer and Hoffrage. *Psychological Review*, *106*, 411–416.
- Lien, Y. & Cheng, P. W. (2000). Distinguishing genuine from spurious causes: A coherence hypothesis. *Cognitive Psychology*, *40*, 87–137.
- Liljeholm, M. & Cheng, P. W. (2007). When is a cause the “same”? Coherent generalization across contexts. *Psychological Science*, *18* (11), 1014–1021.
- Liljeholm, M. & Cheng, P.W. (in press). The influence of virtual sample size on confidence and causal strength judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
- Lober, K. & Shanks, D. R. (2000). Is causal induction based on causal power? Critique of Cheng (1997). *Psychological Review*, *107* (1), 195–212.
- Lu, H., Yuille, A. L., Liljeholm, M., Cheng, P. W., & Holyoak, K. J. (2008). Bayesian generic priors for causal learning. *Psychological Review*, *115*, 955–982.
- Luhmann, C. C. (2005). Confounded: Causal inference and the requirement of independence. In B. G. Bara, L. Barsalou & M. Bucciarelli (Eds.), *Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society* (S. 1355–1360). Mahwah, NJ: Erlbaum.
- Luhmann, C. C. & Ahn, W.-K. (2005). The meaning and computation of causal power: Comment on Cheng (1997) and Novick and Cheng (2004). *Psychological Review*, *112* (3), 685–692.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: Freeman.
- Mayrhofer, R., Goodman, N. D., Waldmann, M. R. & Tenenbaum, J. B. (2008). Structured correlations from the causal background. In B. C. Love, K. McRae & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (S. 303–308). Austin, TX: Cognitive Science Society.

- 
- Meder, B. (2006). Seeing versus doing: Causal Bayes nets as psychological models of causal reasoning. Elektronisch veröffentlichte Dissertation, Georg-August-Universität Göttingen.  
[URL: <http://webdoc.sub.gwdg.de/diss/2006/meder/meder.pdf>]
- Meder, B., Hagmayer, Y. & Waldmann, M. R. (2008). Inferring interventional predictions from observational learning data. *Psychonomic Bulletin & Review*, 15 (1), 75–80.
- Mellers, B. & McGraw, A. P. (1999). How to improve Bayesian reasoning: Comments on Gigerenzer and Hoffrage (1995). *Psychological Review*, 106, 417–424.
- Meltzoff, A. N. (2007). Infants' causal learning. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (S. 37–47). Oxford: Oxford University Press.
- Newsome, G. L. (2003). The debate between current version of covariations and mechanism approaches to causal inference. *Philosophical Psychology*, 16, 87–107.
- Nichols, W. & Danks, D. (2007). Decision making using learned causal structures. In D.S. McNamara & J.G. Trafton (Eds.), *Proceedings of the 29th annual meeting of the cognitive science society* (S. 1343–1348). Austin, TX: Cognitive Science Society.
- Novick, L. R. & Cheng, P. W. (2004). Assessing interactive causal influence. *Psychological Review*, 111 (2), 455–485.
- Oaksford, M. & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. New York: Oxford University Press.
- Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. San Mateo, CA: Morgan Kaufmann Publishers.
- Pearl, J. (2000). *Causation: Models, reasoning and inference*. Cambridge, MA: Cambridge University Press.
- Perales, J. C. & Catena, A. (2006). Human causal induction: A glimpse at the whole picture. *European Journal of Cognitive Psychology*, 18 (2), 277–320.
- Perales, J. C., Catena, A., Maldonado, A., Cándido, A. (2007). The role of mechanism and covariation information in causal belief updating. *Cognition*, 105, 704–714.



- 
- Perales, J. C. & Shanks, D. R. (2007). Models of covariation-based causal judgment: A review and synthesis. *Psychonomic Bulletin & Review*, *14* (4), 577–596.
- Pitman, J. (2002). *Combinatorial stochastic processes. Lecture notes for Saint Flour Summer School.*
- Rehder, B. (2003a). Categorization as causal reasoning. *Cognitive Science*, *27*, 709–748.
- Rehder, B. (2003b). A causal-model theory of conceptual representation and categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 1141–1159.
- Rehder, B. (2006). Human deviations from normative causal reasoning. (S. 2596). In R. Sun (Ed.). *Proceedings of the 28th Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.
- Rehder, B. & Burnett, R. (2005). Feature inference and the causal structure of categories. *Cognitive Psychology*, *50*, 264–314.
- Rehder, B. & Hastie, R. (2001). Causal knowledge and categories: The effects of causal beliefs on categorization, induction, and similarity. *Journal of Experimental Psychology: General*, *130*, 323–360.
- Rehder, B. & Kim, S. (2006). How causal knowledge affects classification: A generative theory of categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32* (4), 659–683.
- Reichenbach, H. (1956). *The direction of time*. Berkeley, CA: University of California Press.
- Rescorla, R. A. & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II. Current research and theory* (S. 64–99). New York: Appleton-Century-Crofts.
- Richardson, T., Schulz, L. & Gopnik, A. (2007). In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (S. 208–230). Oxford: Oxford University Press.
- Roberts, S. & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, *107* (2), 358–367.

- 
- Schulz, L. E., Gopnik, A. & Glymour, C. (2007). Preschool children learn about causal structure from conditional interventions. *Developmental Science*, 10 (3), 322–332.
- Schulz, L., Kushnir, T. & Gopnik, A. (2007). Learning from doing: Intervention and causal inference. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (S. 67–85). Oxford: Oxford University Press.
- Schulz, L. & Sommerville, J. (2006). God does not play dice: Causal determinism and preschoolers' causal inferences. *Child Development*, 77 (2), 427–442.
- Shanks, D. R. (2007). Associationism and cognition: Human contingency learning at 25. *The Quarterly Journal of Experimental Psychology*, 60 (3), 291–309.
- Shultz, T. R. (2007). The Bayesian revolution approaches psychological development. *Developmental Science*, 10 (3), 357–364.
- Sloman, S. A. (2005). *Causal models: How people think about the world and its alternatives*. New York: Oxford University Press.
- Sloman, S. A. & Fernbach, P. M. (2008). The value of rational analysis: An assessment of causal reasoning and learning. In N. Chater & M. Oaksford (Eds.), *The probabilistic mind: Prospects for Bayesian cognitive science* (S. 486–500). New York: Oxford University Press.
- Sloman, S. A. & Hagmayer, Y. (2006). The causal psychology of choice. *Trends in Cognitive Science*, 10 (9), 407–412.
- Sloman, S. A. & Lagnado, D. A. (2005). Do we “do”? *Cognitive Science*, 29 (1), 5–39.
- Sobel, D. M. & Kirkham, N. Z. (2007). Bayes nets and babies: Infants' developing statistical reasoning abilities and their representation of causal knowledge. *Developmental Science*, 10 (3), 298–306.
- Sobel, D. M., Tenenbaum, J. B. & Gopnik, A. (2004). Children's causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers. *Cognitive Science*, 28, 303–333.
- Sober, E. (1987). Discussion: Parsimony, likelihood, and the principle of the common cause. *Philosophy of Science*, 54, 465–469.

- 
- Sommerville, J. A. (2007). Detecting causal structure: The role of interventions in infants' understanding of psychological and physical causal relations. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (S. 48–57). Oxford: Oxford University Press.
- Spellman, B. A. (1996). Acting as intuitive scientists: Contingency judgments are made while controlling for alternative potential causes. *Psychological Science*, 7 (6), 337–342.
- Spirtes, P., Glymour, C. & Scheines, P. (1993, 2000). *Causation, prediction, and search*. New York: Springer.
- Steyvers, M., Griffiths, T. L. & Dennis, S. (2006). Probabilistic inference in human semantic memory. *Trends in Cognitive Science*, 10 (7), 327–334.
- Steyvers, M., Tenenbaum, J. B., Wagenmakers, E. J. & Blum, B. (2003). Inferring causal networks from observations and interventions. *Cognitive Science*, 27, 453–489.
- Tenenbaum, J. B., Griffiths, T. L. & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Science*, 10 (7), 309–318.
- Tenenbaum, J. B., Griffiths, T. L. & Niyogi, S. (2007). Intuitive theories as grammars for causal inference. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (S. 301–322). Oxford: Oxford University Press.
- Waldmann, M. R. (1996). Knowledge-based causal induction. In D. R. Shanks, K. J. Holyoak & D. L. Medin (Eds.), *The psychology of learning and motivation, Vol. 34: Causal learning* (S. 47–88). San Diego, CA, US: Academic Press.
- Waldmann, M. R. (2000). Competition among causes but not effects in predictive and diagnostic learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26 (1), 53–76.
- Waldmann, M. R. (2001). Predictive versus diagnostic causal learning: Evidence from an overshadowing paradigm. *Psychonomic Bulletin & Review*, 8 (3), 600–608.
- Waldmann, M. R. (2007). Combining versus analyzing multiple causes: How domain assumptions and task context affect integration rules. *Cognitive Science*, 31, 233–256.

- 
- Waldmann, M. R., Cheng, P. W., Hagmayer, Y. & Blaisdell, A. P. (2008). Causal learning in rats and humans: A minimal rational model. In N. Chater & M. Oaksford (Eds.), *The probabilistic mind: Prospects for Bayesian cognitive science* (S. 454–484). New York: Oxford University Press.
- Waldmann, M. R. & Hagmayer, Y. (2001). Estimating causal strength: The role of structural knowledge and processing effort. *Cognition*, *82* (1), 27–58.
- Waldmann, M. R. & Hagmayer, Y. (2005). Seeing versus doing: Two modes of accessing causal knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31* (2), 216–227.
- Waldmann, M. R. & Hagmayer, Y. (2006). Categories and causality: The neglected direction. *Cognitive Psychology*, *53*, 27–58.
- Waldmann, M. R., Hagmayer, Y. & Blaisdell, A. P. (2006). Beyond the information given: Causal models in learning and reasoning. *Current Directions in Psychological Science*, *15* (6), 307–311.
- Waldmann, M. R. & Holyoak, K. J. (1992). Predictive and diagnostic learning with causal models: Asymmetries in cue competition. *Journal of Experimental Psychology: General*, *121* (2), 222–236.
- Waldmann, M. R. & Martignon, L. (1998). A Bayesian network model of causal learning. In M. A. Gernsbacher & S. J. Derry (Eds.), *Proceedings of the 20th Annual Conference of the Cognitive Science Society* (S. 1102–1107). Mahwah, NJ: Erlbaum.
- Walsh, C. R. & Sloman, S. A. (2004). Revising causal beliefs. In K. Forbus, D. Gentner & T. Regier (Eds.), *Proceedings of the 26th Annual Conference of the Cognitive Science Society* (S. 1423–1427). Mahwah, NJ: Erlbaum.
- Walsh, C. R. & Sloman, S. A. (2007). Updating beliefs with causal models: Violations of screening off. In M. A. Gluck, J. R. Anderson & S. M. Kosslyn, A *Festschrift for Gordon H. Bower* (S. 345–358). New York: Erlbaum.
- White, P. A. (2002). Causal attribution from covariation information: The evidential evaluation model. *European Journal of Social Psychology*, *32* (5), 667–684.
- White, P. A. (2004). Causal judgment from contingency information: A systematic test of the pCI rule. *Memory and Cognition*, *32*, 353–368.

- 
- White, P. A. (2005). The power PC theory and causal powers: Comment on Cheng (1997) and Novick and Cheng (2004). *Psychological Review*, *112* (3), 675–682.
- White, P. A. (2008). Accounting for occurrence: A new view of the use of contingency information in causal judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34* (1), 204–218.
- Wolff, P. (2007). Representing causation. *Journal of Experimental Psychology: General*, *136* (1), 82–111.
- Wolff, P. & Song, G. (2003). Models of causation and the semantics of causal verbs. *Cognitive Psychology*, *47*, 276–332.
- Woodward, J. (2002). What is a mechanism? A counterfactual account. *Philosophy of Science*, *69*, 8366–8377.
- Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford: Oxford University Press.
- Woodward, J. (2007). Interventionist theories of causation in psychological perspective. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (S. 19–36). Oxford: Oxford University Press.
- Wu, M. & Cheng, P. W. (1999). Why causation need not follow from statistical association: Boundary conditions for the evaluation of generative and preventive causal powers. *Psychological Science*, *10* (2), 92–97.
- Xu, F. & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*, *114* (2), 245–272.

## Anhang A – Varianzanalytische Ergebnisse von Experiment 1

Tabelle A-1

### Varianzanalyse für Experiment 1

Varianzquelle	<i>df</i>	<i>MSE</i>	<i>F</i>	<i>p</i>	$\eta^2$
Zwischen Probanden					
<i>RS</i>	1	14.96	1.56	.218	.03
Fehler	48	9.59			
Innerhalb von Probanden (randomisierte Messwiederholung)					
<i>C</i>	1	918.75	111.85**	<.001	.70
<i>C x RS</i>	1	106.80	13.00**	.001	.21
Fehler	48	8.21			
<i>N<sub>E+</sub></i>	2	54.54	18.21**	<.001	.28
<i>N<sub>E+</sub> x RS</i>	2	25.54	8.53**	<.001	.15
Fehler	96	2.99			
<i>C x N<sub>E+</sub></i>	2	54.31	22.05**	<.001	.31
<i>C x N<sub>E+</sub> x RS</i>	2	7.32	2.97 <sup>+</sup>	.056	.06
Fehler	96	2.46			

Anmerkungen. *RS* = „Reading“-vs.-„Sending“-Faktor; *C* = Status der gemeinsamen Ursache (zwei Stufen: nichts vs. „POR“); *N<sub>E+</sub>* = Anzahl der weiteren an „POR“ denkenden Aliens (drei Stufen: 0, 1 und 2).

<sup>+</sup>*p* < .10, \*\**p* < .01

Tabelle A-2

Konditionale Varianzanalyse für Experiment 1 für Faktor C=0

Varianzquelle	<i>df</i>	<i>MSE</i>	<i>F</i>	<i>P</i>	$\eta^2$
Zwischen Probanden					
<i>RS</i>	1	20.91	1.77	.189	.04
Fehler	48	11.78			
Innerhalb von Probanden (randomisierte Messwiederholung)					
$N_{E+}$	2	3.85	1.52	.223	.03
$N_{E+} \times RS$	2	3.33	1.32	.272	.03
Fehler	96	2.52			

Anmerkungen. *RS* = „Reading“-vs.-„Sending“-Faktor;  $N_{E+}$  = Anzahl der weiteren an „POR“ denkenden Aliens (drei Stufen: 0, 1 und 2).

Tabelle A-3

Konditionale Varianzanalyse für Experiment 1 für Faktor C=1

Varianzquelle	<i>df</i>	<i>MSE</i>	<i>F</i>	<i>P</i>	$\eta^2$
Zwischen Probanden					
<i>RS</i>	1	100.86	16.75**	<.001	.26
Fehler	48	6.02			
Innerhalb von Probanden (randomisierte Messwiederholung)					
$N_{E+}$	2	105.01	35.80**	<.001	.43
$N_{E+} \times RS$	2	29.54	10.07**	<.001	.17
Fehler	96	2.93			

Anmerkungen. *RS* = „Reading“-vs.-„Sending“-Faktor;  $N_{E+}$  = Anzahl der weiteren an „POR“ denkenden Aliens (drei Stufen: 0, 1 und 2).

\*\**p* < .01

## Anhang B – Varianzanalytische Ergebnisse von Experiment 2

Tabelle B-1

### Varianzanalyse für Experiment 2

Varianzquelle	<i>df</i>	<i>MSE</i>	<i>F</i>	<i>p</i>	$\eta^2$
Zwischen Probanden					
<i>RS</i>	1	0.56	0.05	.831	.00
Fehler	48	12.28			
Innerhalb von Probanden (randomisierte Messwiederholung)					
<i>C</i>	1	429.60	85.91**	<.001	.64
<i>C x RS</i>	1	1.20	0.24	.626	.00
Fehler	48	5.00			
-----					
<i>N<sub>E+</sub></i>	2	72.09	19.47**	<.001	.29
<i>N<sub>E+</sub> x RS</i>	2	18.29	4.94**	.009	.09
Fehler	96	3.70			
-----					
<i>C x N<sub>E+</sub></i>	2	11.45	4.60*	.012	.09
<i>C x N<sub>E+</sub> x RS</i>	2	7.89	3.17*	.046	.06
Fehler	96	2.49			

Anmerkungen. *RS* = „Reading“-vs.-„Sending“-Faktor; *C* = Status der gemeinsamen Ursache (zwei Stufen: nichts vs. „POR“); *N<sub>E+</sub>* = Anzahl der weiteren an „POR“ denkenden Aliens (drei Stufen: 0, 1 und 2).

\**p* < .05, \*\**p* < .01



Tabelle B-2

Konditionale Varianzanalyse für Experiment 2 für Faktor C=0

Varianzquelle	<i>df</i>	<i>MSE</i>	<i>F</i>	<i>P</i>	$\eta^2$
Zwischen Probanden					
<i>RS</i>	1	0.06	0.01	.927	<.01
Fehler	48	7.05			
Innerhalb von Probanden (randomisierte Messwiederholung)					
$N_{E+}$	2	13.04	4.34*	.016	.08
$N_{E+} \times RS$	2	2.48	0.83	.441	.02
Fehler	96	3.00			

Anmerkungen. *RS* = „Reading“-vs.-„Sending“-Faktor;  $N_{E+}$  = Anzahl der weiteren an „POR“ denkenden Aliens (drei Stufen: 0, 1 und 2).

\* $p < .05$

Tabelle B-3

Konditionale Varianzanalyse für Experiment 2 für Faktor C=1

Varianzquelle	<i>df</i>	<i>MSE</i>	<i>F</i>	<i>P</i>	$\eta^2$
Zwischen Probanden					
<i>RS</i>	1	1.71	0.17	.685	<.01
Fehler	48	10.23			
Innerhalb von Probanden (randomisierte Messwiederholung)					
$N_{E+}$	2	70.51	22.10**	<.001	.32
$N_{E+} \times RS$	2	23.71	7.43**	.001	.13
Fehler	96	3.19			

Anmerkungen. *RS* = „Reading“-vs.-„Sending“-Faktor;  $N_{E+}$  = Anzahl der weiteren an „POR“ denkenden Aliens (drei Stufen: 0, 1 und 2).

\*\* $p < .01$

## Anhang C – Varianzanalytische Ergebnisse von Experiment 3

Tabelle C-1

### Varianzanalyse für Experiment 3

Varianzquelle	<i>df</i>	<i>MSE</i>	<i>F</i>	<i>p</i>	$\eta^2$
Zwischen Probanden					
<i>RS</i>	1	18.68	2.49	.120	.04
Fehler	58	7.51			
Innerhalb von Probanden (randomisierte Messwiederholung)					
<i>C</i>	1	1432.01	208.92**	<.001	.78
<i>C x RS</i>	1	12.10	1.77	.189	.03
Fehler	58	6.85			
<i>N<sub>E+</sub></i>	2	38.39	14.69**	<.001	.20
<i>N<sub>E+</sub> x RS</i>	2	15.09	5.77**	.004	.09
Fehler	116	2.61			
<i>C x N<sub>E+</sub></i>	2	7.75	3.92*	.022	.06
<i>C x N<sub>E+</sub> x RS</i>	2	6.31	3.19*	.045	.05
Fehler	116	1.98			

Anmerkungen. *RS* = „Reading“-vs.-„Sending“-Faktor; *C* = Status der gemeinsamen Ursache (zwei Stufen: nichts vs. „POR“); *N<sub>E+</sub>* = Anzahl der weiteren an „POR“ denkenden Aliens (drei Stufen: 0, 1 und 2).

\**p* < .05, \*\**p* < .01

Tabelle C-2

Konditionale Varianzanalyse für Experiment 3 für Faktor C=0

Varianzquelle	<i>df</i>	<i>MSE</i>	<i>F</i>	<i>P</i>	$\eta^2$
Zwischen Probanden					
<i>RS</i>	1	0.36	0.05	.824	<.01
Fehler	58	7.13			
Innerhalb von Probanden (randomisierte Messwiederholung)					
$N_{E+}$	2	6.87	4.93**	.009	.08
$N_{E+} \times RS$	2	0.97	0.70	.500	.01
Fehler	116	1.39			

Anmerkungen. *RS* = „Reading“-vs.-„Sending“-Faktor;  $N_{E+}$  = Anzahl der weiteren an „POR“ denkenden Aliens (drei Stufen: 0, 1 und 2).

\*\**p* < .01

Tabelle C-3

Konditionale Varianzanalyse für Experiment 3 für Faktor C=1

Varianzquelle	<i>df</i>	<i>MSE</i>	<i>F</i>	<i>P</i>	$\eta^2$
Zwischen Probanden					
<i>RS</i>	1	30.42	4.20*	.045	.07
Fehler	58	7.24			
Innerhalb von Probanden (randomisierte Messwiederholung)					
$N_{E+}$	2	39.27	12.29**	<.001	.17
$N_{E+} \times RS$	2	20.42	6.39**	.002	.10
Fehler	116	3.20			

Anmerkungen. *RS* = „Reading“-vs.-„Sending“-Faktor;  $N_{E+}$  = Anzahl der weiteren an „POR“ denkenden Aliens (drei Stufen: 0, 1 und 2).

\**p* < .05, \*\**p* < .01

## Anhang D – Ausführliche Ableitung des Basismodells

Es sei  $C \in \{0,1\}$  eine binäre Ursache und  $E_1, \dots, E_n$  mit  $E_i \in \{0,1\} \forall i \in \{1, \dots, n\}$  binäre Effekte dieser Ursache (*Common-Cause*-Struktur),  $PN \in \{0,1\}$  eine unbeobachtete präventive Ursache und  $w = (w_C, w_{PN}, b_C, b_{PN}, b_E)$  der zugehörige Parametervektor mit

$w_C \in [0,1]$  – die Kausalstärke von  $C$  im Hinblick auf die Effekte,

$w_{PN} \in [0,1]$  – die präventive Stärke von  $PN$  im Hinblick auf die Links  
 $C-E_i$ ,

$b_C \in [0,1]$  – die Basisrate der gemeinsamen Ursache  $C$ ,

$b_{PN} \in [0,1]$  – die Basisrate von  $PN$  (*preventive noise*) und

$b_E \in [0,1]$  – die Basisrate der Effekte  $E_1, \dots, E_n$ .

Es gelte ferner für die *A-priori*-Verteilungen dieser Parameter

$$w_C \sim \text{Beta}(\alpha_{w_C}, \beta_{w_C}) \text{ mit Erwartungswert } E[w_C] = \frac{\alpha_{w_C}}{\alpha_{w_C} + \beta_{w_C}}$$

$$w_{PN} \sim \text{Beta}(\alpha_{w_{PN}}, \beta_{w_{PN}}) \text{ mit Erwartungswert } E[w_{PN}] = \frac{\alpha_{w_{PN}}}{\alpha_{w_{PN}} + \beta_{w_{PN}}}$$

$$b_C \sim \text{Beta}(\alpha_{b_C}, \beta_{b_C}) \text{ mit Erwartungswert } E[b_C] = \frac{\alpha_{b_C}}{\alpha_{b_C} + \beta_{b_C}}$$

$$b_{PN} \sim \text{Beta}(\alpha_{b_{PN}}, \beta_{b_{PN}}) \text{ mit Erwartungswert } E[b_{PN}] = \frac{\alpha_{b_{PN}}}{\alpha_{b_{PN}} + \beta_{b_{PN}}}$$

$$b_E \sim \text{Beta}(\alpha_{b_E}, \beta_{b_E}) \text{ mit Erwartungswert } E[b_E] = \frac{\alpha_{b_E}}{\alpha_{b_E} + \beta_{b_E}}$$

und sei damit  $\omega$  der Vektor der Hyperparameter (der das gesamte abstrakte Vorwissen beschreibt) mit

$$\omega = (\alpha_{w_C}, \beta_{w_C}, \alpha_{w_{PN}}, \beta_{w_{PN}}, \alpha_{b_C}, \beta_{b_C}, \alpha_{b_{PN}}, \beta_{b_{PN}}, \alpha_{b_E}, \beta_{b_E}) \quad (\text{D-1})$$

Für die Verknüpfung von  $C$  und  $PN$  im Hinblick auf  $E$  gelte ein *Noisy-AND-NOT*-Gate mit anschließendem *Noisy-OR-Gate* (siehe u.a. Glymour, 2001) mit folgender bedingter Wahrscheinlichkeitstabelle:

$C$	$PN$	$P(E_i = 1   C, PN)$
0	0	$b_E$
0	1	$b_E$
1	0	$w_C + b_E - w_C b_E$
1	1	$w_C(1 - w_{PN}) + b_E - w_C(1 - w_{PN})b_E$

Seien Ursache  $C$  sowie die Effekte  $E_1, \dots, E_{n-1}$  beobachtet, Effekt  $E_n$  jedoch unbeobachtet (Zieleffekt). Dann gilt für die Wahrscheinlichkeit der Anwesenheit des Zieleffekts ( $E_n = 1$ ) gegeben der beobachteten Variablen:

$$P(E_n = 1|C, E_1, \dots, E_{n-1}; \omega) = \iint P(E_n = 1|C, E_1, \dots, E_{n-1}; w) \cdot P(w|\omega) dw \quad (D-2)$$

unter der Annahme, dass<sup>63</sup>

$$P(w|C, E_1, \dots, E_{n-1}; \omega) = P(w|\omega) \quad (D-3)$$

Ferner folgt für

$$\begin{aligned} P(E_n = 1|C, E_1, \dots, E_{n-1}; w) \\ = \sum_{PN} P(E_n = 1|C, PN, E_1, \dots, E_{n-1}; w) \cdot P(PN|C, E_1, \dots, E_{n-1}; w) \end{aligned} \quad (D-4)$$

$$= \sum_{PN} P(E_n = 1|C, PN; w) \cdot P(PN|C, E_1, \dots, E_{n-1}; w) \quad (D-5)$$

$$= \sum_{PN} P(E_n = 1|C, PN; w) \cdot \frac{P(E_1, \dots, E_{n-1}|C, PN; w)P(PN|C; w)}{P(E_1, \dots, E_{n-1}|C; w)} \quad (D-6)$$

$$= \sum_{PN} P(E_n = 1|C, PN; w) \cdot \frac{P(E_1, \dots, E_{n-1}|C, PN; w)P(PN|w)}{\sum_{PN} P(E_1, \dots, E_{n-1}|C, PN; w)P(PN|w)} \quad (D-7)$$

$$= \frac{\sum_{PN} P(E_n = 1|C, PN; w) \cdot P(E_1, \dots, E_{n-1}|C, PN; w)P(PN|w)}{\sum_{PN} P(E_1, \dots, E_{n-1}|C, PN; w)P(PN|w)} \quad (D-8)$$

Sei nun  $m$  die Anzahl der als anwesend beobachteten Effekte und folglich damit  $(n - m - 1)$  die Anzahl der als abwesend beobachteten Effekte. Da aufgrund der Markov-Bedingung gilt

$$P(E_1, \dots, E_{n-1}|PN, C; w) = P(E_1|C, PN; w) \cdot \dots \cdot P(E_{n-1}|C, PN; w) \quad (D-9)$$

folgt

$$\begin{aligned} P(E_1, \dots, E_{n-1}|PN, C; w) \\ = P(E = 1|C, PN; w)^m \cdot P(E = 0|C, PN; w)^{n-m-1} \end{aligned} \quad (D-10)$$

$$\text{mit } P(E = 0|C, PN; w) = 1 - P(E = 1|C, PN; w) \quad (D-11)$$

<sup>63</sup> Zur Begründung dieser Annahme siehe 4.2.

Damit gilt vereinfachend für Gleichung D-8:

$$\begin{aligned}
 & P(E_n = 1|C, E_1, \dots, E_{n-1}; w) \\
 &= \frac{\sum_{PN} P(E = 1|C, PN; w)^{m+1} \cdot P(E = 0|C, PN; w)^{n-m-1} \cdot P(PN|w)}{\sum_{PN} P(E = 1|C, PN; w)^m \cdot P(E = 0|C, PN; w)^{n-m-1} \cdot P(PN|w)} \quad (D-12)
 \end{aligned}$$

Gleichung D-12 kann nun eingesetzt werden in Gleichung D-2. Es folgt:

$$\begin{aligned}
 & P(E_n = 1|C, E_1, \dots, E_{n-1}; \omega) \\
 &= \iint \frac{\sum_{PN} P(E = 1|C, PN; w)^{m+1} \cdot P(E = 0|C, PN; w)^{n-m-1} \cdot P(PN|w)}{\sum_{PN} P(E = 1|C, PN; w)^m \cdot P(E = 0|C, PN; w)^{n-m-1} \cdot P(PN|w)} \cdot P(w|\omega) dw \quad (D-13)
 \end{aligned}$$

Da das Mehrfachintegral analytisch nicht lösbar ist, wird die Vorhersage mit Hilfe einer Monte-Carlo-Simulation approximiert. Dazu sei

$$w^{(j)} = (w_C^{(j)}, w_{PN}^{(j)}, b_C^{(j)}, b_{PN}^{(j)}, b_E^{(j)})$$

ein Vektor mit Parameterwerten unabhängig gezogen aus den jeweiligen *A-priori*-Verteilungen der Parameter. Dies werde  $M$ -mal wiederholt mit  $j = 1, \dots, M$ . Aufbauend auf Gleichung D-13 folgt (siehe dazu auch Griffiths & Tenenbaum, 2005):

$$\begin{aligned}
 & P(E_n = 1|C, E_1, \dots, E_{n-1}; \omega) \\
 &= \frac{1}{M} \sum_{j=1}^M \frac{\sum_{PN} P(E = 1|C, PN; w^{(j)})^{m+1} \cdot P(E = 0|C, PN; w^{(j)})^{n-m-1} \cdot P(PN|w^{(j)})}{\sum_{PN} P(E = 1|C, PN; w^{(j)})^m \cdot P(E = 0|C, PN; w^{(j)})^{n-m-1} \cdot P(PN|w^{(j)})} \quad (D-14)
 \end{aligned}$$

Dies lässt sich einfach bestimmen, da

$$\begin{aligned}
 & P(E = 1|C, PN; w^{(j)}) = 1 - P(E = 0|C, PN; w^{(j)}) \\
 &= \begin{cases} b_E^{(j)} & |C = 0, PN = 0 \\ b_E^{(j)} & |C = 0, PN = 1 \\ w_C^{(j)} + b_E^{(j)} - w_C^{(j)} b_E^{(j)} & |C = 1, PN = 0 \\ w_C^{(j)} (1 - w_{PN}^{(j)}) + b_E^{(j)} - w_C^{(j)} (1 - w_{PN}^{(j)}) b_E^{(j)} & |C = 1, PN = 1 \end{cases} \quad (D-15)
 \end{aligned}$$

und

$$P(PN|w^{(j)}) = \begin{cases} 1 - b_{PN}^{(j)} & |PN = 0 \\ b_{PN}^{(j)} & |PN = 1 \end{cases} \quad (D-16)$$

## Anhang E – Varianzanalytische Ergebnisse von Experiment 4

Tabelle E-1

### Varianzanalyse für Experiment 4

Varianzquelle	<i>df</i>	<i>MSE</i>	<i>F</i>	<i>p</i>	$\eta^2$
Zwischen Probanden					
<i>RS</i>	1	16.30	6.26*	.015	.10
Error	54	2.60			
Innerhalb von Probanden (randomisierte Messwiederholung)					
<i>C</i>	1	1107.44	146.05**	<.001	.73
<i>C x RS</i>	1	29.76	3.93 <sup>+</sup>	.053	.07
Fehler	54	7.58			
-----					
$N_{E+}$	2	161.58	40.39**	<.001	.43
$N_{E+} \times RS$	2	42.07	10.52**	<.001	.16
Fehler	108	4.00			
-----					
$C \times N_{E+}$	2	12.91	5.19**	.007	.09
$C \times N_{E+} \times RS$	2	3.40	1.37	.260	.02
Fehler	108	2.49			

Anmerkungen. *RS* = „Reading“-vs.-„Sending“-Faktor; *C* = Status der gemeinsamen Ursache (zwei Stufen: „TUS“ vs. „POR“);  $N_{E+}$  = Anzahl der weiteren an „POR“ denkenden Aliens (drei Stufen: 0, 1 und 2).

\* $p < .05$ , \*\* $p < .01$

Tabelle E-2

Konditionale Varianzanalyse für Experiment 4 für Faktor C=„A“ („TUS“)

Varianzquelle	<i>df</i>	<i>MSE</i>	<i>F</i>	<i>P</i>	$\eta^2$
Zwischen Probanden					
<i>RS</i>	1	1.01	0.25	.619	<.01
Fehler	54	4.01			
Innerhalb von Probanden (randomisierte Messwiederholung)					
$N_{E+}$	2	53.61	20.25**	<.001	.27
$N_{E+} \times RS$	2	11.11	4.20*	.018	.07
Fehler	108	2.65			

Anmerkungen. *RS* = „Reading“-vs.-„Sending“-Faktor;  $N_{E+}$  = Anzahl der weiteren an „POR“ denkenden Aliens (drei Stufen: 0, 1 und 2).

\* $p < .05$ , \*\* $p < .01$

Tabelle E-3

Konditionale Varianzanalyse für Experiment 4 für Faktor C=„B“ („POR“)

Varianzquelle	<i>df</i>	<i>MSE</i>	<i>F</i>	<i>P</i>	$\eta^2$
Zwischen Probanden					
<i>RS</i>	1	45.05	7.30**	.009	.12
Fehler	54	6.17			
Innerhalb von Probanden (randomisierte Messwiederholung)					
$N_{E+}$	2	120.88	31.47**	<.001	.37
$N_{E+} \times RS$	2	34.36	8.94**	<.001	.14
Fehler	108	3.84			

Anmerkungen. *RS* = „Reading“-vs.-„Sending“-Faktor;  $N_{E+}$  = Anzahl der weiteren an „POR“ denkenden Aliens (drei Stufen: 0, 1 und 2).

\*\* $p < .01$



## Anhang F – Herleitung des Modells für *Causal-Chain*-Strukturen

Es seien  $C_1, \dots, C_n \in \{0,1\}$   $n$  binäre Variablen dergestalt, dass  $C_i \rightarrow C_{i+1} \forall i \in \{1, \dots, n-1\}$  gilt und  $E \in \{0,1\}$  binärer Effekt von Variable  $C_n$ , also  $C_n \rightarrow E$ , ist (*Causal-Chain*-Struktur).<sup>64</sup> Des Weiteren seien  $PN_{C_1}, \dots, PN_{C_n} \in \{0,1\}$   $n$  unbeobachtete präventive Ursachen (siehe dazu Abbildung 24) und  $w = (w_C, w_{PN}, b_C, b_{PN}, b_E)$  der zum System gehörende Parametervektor mit

$w_C \in [0,1]$  – die Kausalstärke aller Relationen  $X \rightarrow Y$ ,

$w_{PN} \in [0,1]$  – die präventive Stärke der Ursachen  $PN_{C_1}, \dots, PN_{C_n}$

$b_C \in [0,1]$  – die Basisrate der Ursachen  $C_1, \dots, C_n$

$b_{PN} \in [0,1]$  – die Basisrate der präventiven Ursachen  $PN_{C_1}, \dots, PN_{C_n}$

$b_E \in [0,1]$  – die Basisrate des Effekts  $E$

Für die *A-priori*-Verteilungen dieser Parameter siehe Anhang D entsprechend.

Für die Verknüpfung von  $C_n$  und  $PN_{C_n}$  im Hinblick auf  $E$  gelte ein *Noisy-AND-NOT*-Gate mit anschließendem *Noisy-OR-Gate* (siehe u.a. Glymour, 2001) mit folgender bedingter Wahrscheinlichkeitstabelle:<sup>65</sup>

$C_n$	$PN_{C_n}$	$P(E = 1   C_n, PN_{C_n})$
0	0	$b_E$
0	1	$b_E$
1	0	$w_C + b_E - w_C b_E$
1	1	$w_C(1 - w_{PN}) + b_E - w_C(1 - w_{PN})b_E$

Seien die Variablen  $C_1, \dots, C_n$  beobachtet, der ultimative Effekt  $E$  (Zieleffekt) sowie die präventiven Ursachen  $PN_{C_1}, \dots, PN_{C_n}$  unbeobachtet. Dann gilt für die Wahrscheinlichkeit der Anwesenheit des Zieleffekts ( $E = 1$ ) gegeben der beobachteten Variablen:

$$P(E = 1 | C_1, \dots, C_n; \omega) = \iint P(E = 1 | C_1, \dots, C_n; w) \cdot P(w | \omega) dw \quad (\text{F-1})$$

unter der Annahme, dass<sup>66</sup>

<sup>64</sup> Die Ableitung erfolgt allgemeiner als in Abschnitt 5.2.1, in dem die Ableitung für  $n=3$  dargestellt ist. Es gilt mithin:  $C_1 = IC_1, C_2 = IC_2, C_3 = DC$ .

<sup>65</sup> Für die Relationen  $C_i \rightarrow C_{i+1}$  gelte die Wahrscheinlichkeitstabelle entsprechend.

<sup>66</sup> Zur Begründung dieser Annahme siehe Abschnitt 4.2.

$$P(w|C_1, \dots, C_n; \omega) = P(w|\omega) \quad (\text{F-2})$$

Ferner folgt unter Einführung von  $PN_{C_n}$  (da  $E$  im um die Fehlerquellen erweiterten Netzwerk nur von  $C_n$  und  $PN_{C_n}$  abhängig ist) für

$$\begin{aligned} P(E = 1|C_1, \dots, C_n; w) \\ = P(E = 1|C_n; w) \end{aligned} \quad (\text{F-3})$$

$$= \sum_{PN_{C_n}} P(E = 1|C_n, PN_{C_n}; w) \cdot P(PN_{C_n}; w) \quad (\text{F-4})$$

$$\begin{aligned} = P(E = 1|C_n, PN_{C_n} = 0; w) \cdot P(PN_{C_n} = 0; w) + \\ P(E = 1|C_n, PN_{C_n} = 1; w) \cdot P(PN_{C_n} = 1; w) \end{aligned} \quad (\text{F-5})$$

Für  $C_n = 0$  folgt:

$$P(E = 1|C_n = 0; w) = b_E(1 - b_{PN}) + b_E b_{PN} = b_E \quad (\text{F-6})$$

Für  $C_n = 1$  folgt:

$$\begin{aligned} P(E = 1|C_n = 1; w) \\ = (b_E + w_C - b_E w_C)(1 - b_{PN}) + (b_E + w_C(1 - w_{PN}) \\ - b_E w_C(1 - w_{PN}))b_{PN} \end{aligned} \quad (\text{F-7})$$

$$\begin{aligned} = b_E + w_C - b_E - b_{PN} w_C + b_{PN} b_E w_C + b_{PN} w_C(1 - w_{PN}) - b_{PN} b_E w_C(1 \\ - w_{PN}) \end{aligned} \quad (\text{F-8})$$

$$= b_E + w_C(1 - b_{PN} + b_{PN}(1 - w_{PN})) - b_E w_C(1 - b_{PN} + b_{PN}(1 - w_{PN})) \quad (\text{F-9})$$

$$= b_E + w_C(1 - b_{PN} w_{PN}) - b_E w_C(1 - b_{PN} w_{PN}) \quad (\text{F-10})$$

$$= b_E + w_C^* - b_E w_C^* \quad (\text{F-11})$$

mit

$$w_C^* = w_C(1 - b_{PN} w_{PN}) \quad (\text{F-12})$$

Damit entspricht die Inferenz über den Zieleffekt  $E$  in einer Struktur mit unbeobachteten Fehlerursachen der Inferenz in einem Bayes-Netz ohne diese Komponenten mit angepasster Kausalstärke. Einen Einfluss dieser Komponenten über die Verringerung der beobachteten Kausalstärke hinaus gibt es daher nicht.

## Anhang G – Varianzanalytische Ergebnisse – Experiment 5

Tabelle G-1

### Varianzanalyse für Experiment 5

Varianzquelle	<i>df</i>	<i>MSE</i>	<i>F</i>	<i>p</i>	$\eta^2$
Zwischen Probanden					
<i>RS</i>	1	2.08	0.33	.571	.01
Fehler	48	6.39			
Innerhalb von Probanden (randomisierte Messwiederholung)					
<i>DC</i>	1	1564.08	191.99**	<.001	.80
<i>DC x RS</i>	1	1.20	0.15	.702	.00
Fehler	48	8.15			
-----					
<i>N<sub>IC+</sub></i>	2	11.59	15.31**	<.001	.24
<i>N<sub>IC+</sub> x RS</i>	2	0.74	0.98	.378	.02
Fehler	96	0.76			
-----					
<i>DC x N<sub>IC+</sub></i>	2	0.54	0.86	.426	.02
<i>DC x N<sub>IC+</sub> x RS</i>	2	0.00	0.01	.995	.00
Fehler	96	0.63			

Anmerkungen. *RS* = „Reading“-vs.-„Sending“-Faktor; *DC* = Status der direkten Ursache (zwei Stufen: nichts vs. „POR“); *N<sub>IC+</sub>* = Anzahl der als anwesend beobachteten indirekten Ursachen (drei Stufen: 0, 1 und 2).

\*\**p* < .01

Tabelle G-2

Konditionale Varianzanalyse für Experiment 5 für Faktor DC=0 (nichts)

Varianzquelle	df	MSE	F	P	$\eta^2$
Zwischen Probanden					
RS	1	3.23	0.89	.349	.02
Fehler	48	3.61			
Innerhalb von Probanden (randomisierte Messwiederholung)					
$N_{IC+}$	2	5.13	6.47**	.002	.12
$N_{IC+} \times RS$	2	0.33	0.41	.663	.01
Fehler	96	0.79			

Anmerkungen. RS = „Reading“-vs.-„Sending“-Faktor;  $N_{IC+}$  = Anzahl der als anwesend beobachteten indirekten Ursachen (drei Stufen: 0, 1 und 2).

\*\*p < .01

Tabelle G-3

Konditionale Varianzanalyse für Experiment 5 für Faktor DC=1 („POR“)

Varianzquelle	df	MSE	F	P	$\eta^2$
Zwischen Probanden					
RS	1	0.06	0.01	.941	<.01
Fehler	48	10.93			
Innerhalb von Probanden (randomisierte Messwiederholung)					
$N_{IC+}$	2	7.01	11.77**	<.001	.20
$N_{IC+} \times RS$	2	0.42	0.71	.496	.01
Fehler	96	0.60			

Anmerkungen. RS = „Reading“-vs.-„Sending“-Faktor;  $N_{IC+}$  = Anzahl der als anwesend beobachteten indirekten Ursachen (drei Stufen: 0, 1 und 2).

\*\*p < .01

## Anhang H – Herleitung für den Fall weiterer Fehlerquellen

Es sei  $C \in \{0,1\}$  eine binäre Ursache und  $E_1, \dots, E_n$  mit  $E_i \in \{0,1\} \forall i \in \{1, \dots, n\}$  binäre Effekte dieser Ursache (*Common-Cause-Struktur*),  $PN \in \{0,1\}$  eine unbeobachtete präventive Ursache für alle Relationen  $C-E_i$ ,  $PN^* \in \{0,1\}$  eine beobachtete präventive Ursache für alle Relationen  $C-E_i$  außer  $C-E_n$  und  $w = (w_C, w_{PN}, w_{PN^*}, b_C, b_{PN}, b_{PN^*}, b_E)$  der zugehörige Parametervektor mit

$w_C \in [0,1]$  – die Kausalstärke von  $C$  im Hinblick auf die Effekte,

$w_{PN} \in [0,1]$  – die präventive Stärke von  $PN$  im Hinblick auf die Links  $C-E_i$ ,

$w_{PN^*} \in [0,1]$  – die präventive Stärke von  $PN^*$  im Hinblick auf die Links  $C-E_i$  (außer  $C-E_n$ ),

$b_C \in [0,1]$  – die Basisrate der gemeinsamen Ursache  $C$ ,

$b_{PN} \in [0,1]$  – die Basisrate von  $PN$  (*preventive noise*) und

$b_{PN^*} \in [0,1]$  – die Basisrate von  $PN^*$  und

$b_E \in [0,1]$  – die Basisrate der Effekte  $E_1, \dots, E_n$ .

Für die *A-priori*-Verteilungen dieser Parameter siehe Anhang D entsprechend.

Für die Verknüpfung von  $C$  und  $PN$  im Hinblick auf den Effekt  $E_n$  gelte die bedingte Wahrscheinlichkeitstabelle aus Anhang D (da  $PN^*$  nicht auf  $E_n$  wirkt). Für alle anderen Effekte  $E_1, \dots, E_{n-1}$ , gelte ebenfalls ein *Noisy-AND-NOT*-Gate (mit zwei präventiven Ursachen) mit anschließendem *Noisy-OR-Gate* (siehe u.a. Glymour, 2001) mit folgender bedingter Wahrscheinlichkeitstabelle:

$C$	$PN$	$PN^*$	$P(E_i = 1 C, PN, PN^*)$ für $i = 1, \dots, n - 1$
0	0	0	$b_E$
0	0	1	$b_E$
0	1	0	$b_E$
0	1	1	$b_E$
1	0	0	$w_C + b_E - w_C b_E$
1	0	1	$w_C(1 - w_{PN^*}) + b_E - w_C(1 - w_{PN^*})b_E$
1	1	0	$w_C(1 - w_{PN}) + b_E - w_C(1 - w_{PN})b_E$
1	1	1	$w_C(1 - w_{PN})(1 - w_{PN^*}) + b_E - w_C(1 - w_{PN})(1 - w_{PN^*})b_E$

Seien Ursache  $C$ , die Effekte  $E_1, \dots, E_{n-1}$  sowie  $PN^*$  beobachtet, Effekt  $E_n$  (und die Fehlerursache  $PN$ ) jedoch unbeobachtet (Zieleffekt). Dann gilt für die Wahrscheinlichkeit der Anwesenheit des Zieleffekts ( $E_n = 1$ ) gegeben der beobachteten Variablen analog zu Anhang D:

$$\begin{aligned} P(E_n = 1|C, PN^*, E_1, \dots, E_{n-1}; \omega) \\ = \iint P(E_n = 1|C, PN^*, E_1, \dots, E_{n-1}; w) \cdot P(w|\omega) dw \end{aligned} \quad (H-1)$$

unter der Annahme, dass<sup>67</sup>

$$P(w|C, E_1, \dots, E_{n-1}; \omega) = P(w|\omega) \quad (H-2)$$

Ferner folgt für

$$\begin{aligned} P(E_n = 1|C, PN^*, E_1, \dots, E_{n-1}; w) \\ = \sum_{PN} P(E_n = 1|C, PN^*, PN, E_1, \dots, E_{n-1}; w) \cdot P(PN|C, PN^*, E_1, \dots, E_{n-1}; w) \end{aligned} \quad (H-3)$$

$$= \sum_{PN} P(E_n = 1|C, PN; w) \cdot P(PN|C, PN^*, E_1, \dots, E_{n-1}; w) \quad (H-4)$$

$$= \sum_{PN} P(E_n = 1|C, PN; w) \cdot \frac{P(E_1, \dots, E_{n-1}|C, PN, PN^*; w)P(PN|C, PN^*; w)}{P(E_1, \dots, E_{n-1}|C, PN^*; w)} \quad (H-5)$$

$$= \sum_{PN} P(E_n = 1|C, PN; w) \cdot \frac{P(E_1, \dots, E_{n-1}|C, PN, PN^*; w)P(PN|w)}{\sum_{PN} P(E_1, \dots, E_{n-1}|C, PN, PN^*; w)P(PN|w)} \quad (H-6)$$

$$= \frac{\sum_{PN} P(E_n = 1|C, PN; w) \cdot P(E_1, \dots, E_{n-1}|C, PN, PN^*; w)P(PN|w)}{\sum_{PN} P(E_1, \dots, E_{n-1}|C, PN, PN^*; w)P(PN|w)} \quad (H-7)$$

Sei nun  $m$  die Anzahl der als anwesend beobachteten Effekte und folglich damit  $(n - m - 1)$  die Anzahl der als abwesend beobachteten Effekte. Da aufgrund der Markov-Bedingung gilt

$$\begin{aligned} P(E_1, \dots, E_{n-1}|C, PN, PN^*; w) \\ = P(E_1|C, PN, PN^*; w) \cdot \dots \cdot P(E_{n-1}|C, PN, PN^*; w) \end{aligned} \quad (H-8)$$

folgt

<sup>67</sup> Zur Begründung dieser Annahme siehe 4.2.

$$P(E_1, \dots, E_{n-1} | C, PN, PN^*; w) = P(E = 1 | C, PN, PN^*; w)^m \cdot P(E = 0 | C, PN, PN^*; w)^{n-m-1} \quad (\text{H-9})$$

$$\text{mit } P(E = 0 | C, PN, PN^*; w) = 1 - P(E = 1 | C, PN, PN^*; w) \quad (\text{H-10})$$

Damit gilt vereinfachend für Gleichung H-7:

$$P(E_n = 1 | C, PN^*, E_1, \dots, E_{n-1}; w) = \frac{\sum_{PN} P(E_n = 1 | C, PN; w) \cdot P(E = 1 | C, PN, PN^*; w)^m \cdot P(E = 0 | C, PN, PN^*; w)^{n-m-1} \cdot P(PN | w)}{\sum_{PN} P(E = 1 | C, PN, PN^*; w)^m \cdot P(E = 0 | C, PN, PN^*; w)^{n-m-1} \cdot P(PN | w)} \quad (\text{H-11})$$

Gleichung H-11 kann nun eingesetzt werden in Gleichung H-1. Es folgt:

$$P(E_n = 1 | C, E_1, \dots, E_{n-1}; \omega) = \iint \frac{\sum_{PN} P(E_n = 1 | C, PN; w) \cdot P(E = 1 | C, PN, PN^*; w)^m \cdot P(E = 0 | C, PN, PN^*; w)^{n-m-1} \cdot P(PN | w)}{\sum_{PN} P(E = 1 | C, PN, PN^*; w)^m \cdot P(E = 0 | C, PN, PN^*; w)^{n-m-1} \cdot P(PN | w)} \cdot P(w | \omega) dw \quad (\text{H-12})$$

Da das Mehrfachintegral analytisch nicht lösbar ist, wird die Vorhersage mit Hilfe einer Monte-Carlo-Simulation approximiert. Dazu sei

$$w^{(j)} = (w_C^{(j)}, w_{PN}^{(j)}, w_{PN^*}^{(j)}, b_C^{(j)}, b_{PN}^{(j)}, b_{PN^*}^{(j)}, b_E^{(j)})$$

ein Vektor mit Parameterwerten unabhängig gezogen aus den jeweiligen *A-priori*-Verteilungen der Parameter. Dies werde M-mal wiederholt mit  $j = 1, \dots, M$ . Aufbauend auf Gleichung H-12 folgt (siehe dazu auch Griffiths & Tenenbaum, 2005):

$$P(E_n = 1 | C, PN^*, E_1, \dots, E_{n-1}; \omega) = \frac{1}{M} \sum_{j=1}^M \frac{\sum_{PN} P(E_n = 1 | C, PN; w) \cdot P(E = 1 | C, PN, PN^*; w)^m \cdot P(E = 0 | C, PN, PN^*; w)^{n-m-1} \cdot P(PN | w)}{\sum_{PN} P(E = 1 | C, PN, PN^*; w)^m \cdot P(E = 0 | C, PN, PN^*; w)^{n-m-1} \cdot P(PN | w)} \quad (\text{H-13})$$

Dies lässt sich einfach bestimmen, da

$$P(E_n = 1 | C, PN; w^{(j)}) = 1 - P(E_n = 0 | C, PN; w^{(j)}) = \begin{cases} b_E^{(j)} & | C = 0, PN = 0 \\ b_E^{(j)} & | C = 0, PN = 1 \\ w_C^{(j)} + b_E^{(j)} - w_C^{(j)} b_E^{(j)} & | C = 1, PN = 0 \\ w_C^{(j)} (1 - w_{PN}^{(j)}) + b_E^{(j)} - w_C^{(j)} (1 - w_{PN}^{(j)}) b_E^{(j)} & | C = 1, PN = 1 \end{cases} \quad (\text{H-14})$$

und

$$P(E = 1|C, PN, PN^*; w^{(j)}) = 1 - P(E = 0|C, PN, PN^*; w^{(j)})$$

$$= \begin{cases} b_E^{(j)} & |C = 0, PN = 0, PN^* = 0 \\ b_E^{(j)} & |C = 0, PN = 0, PN^* = 1 \\ b_E^{(j)} & |C = 0, PN = 1, PN^* = 0 \\ b_E^{(j)} & |C = 0, PN = 1, PN^* = 1 \\ w_C^{(j)} + b_E^{(j)} - w_C^{(j)}b_E^{(j)} & |C = 1, PN = 0, PN^* = 0 \\ w_C^{(j)}(1 - w_{PN^*}^{(j)}) + b_E^{(j)} - w_C^{(j)}(1 - w_{PN^*}^{(j)})b_E^{(j)} & |C = 1, PN = 0, PN^* = 1 \\ w_C^{(j)}(1 - w_{PN}^{(j)}) + b_E^{(j)} - w_C^{(j)}(1 - w_{PN}^{(j)})b_E^{(j)} & |C = 1, PN = 1, PN^* = 0 \\ w_C^{(j)}(1 - w_{PN}^{(j)})(1 - w_{PN^*}^{(j)}) + b_E^{(j)} - w_C^{(j)}(1 - w_{PN}^{(j)})(1 - w_{PN^*}^{(j)})b_E^{(j)} & |C = 1, PN = 1, PN^* = 1 \end{cases} \quad (\text{H-15})$$

sowie

$$P(PN|w^{(j)}) = \begin{cases} 1 - b_{PN}^{(j)} & |PN = 0 \\ b_{PN}^{(j)} & |PN = 1 \end{cases} \quad (\text{H-16})$$



## Anhang I – Varianzanalytische Ergebnisse – Experiment 6

Tabelle I-1

Varianzanalyse für Experiment 6 (nur Kontrollbedingungen)

Varianzquelle	<i>df</i>	<i>MSE</i>	<i>F</i>	<i>p</i>	$\eta^2$
Zwischen Probanden					
<i>EXT<sub>Control</sub></i>	1	0.02	0.00	.959	.00
Error	30	7.92			
Innerhalb von Probanden (randomisierte Messwiederholung)					
<i>C</i>	1	462.52	71.91**	<.001	.71
<i>C x EXT<sub>Control</sub></i>	1	0.19	0.03	.866	.00
Error	30	6.43			
-----					
<i>N<sub>E+</sub></i>	2	33.19	10.57**	<.001	.26
<i>N<sub>E+</sub> x EXT<sub>Control</sub></i>	2	0.57	0.18	.835	.01
Error	60	3.14			
-----					
<i>C x N<sub>E+</sub></i>	2	34.54	14.69**	<.001	.33
<i>C x N<sub>E+</sub> x EXT<sub>Control</sub></i>	2	0.58	0.25	.783	.01
Error	60	2.35			

Anmerkungen. *EXT<sub>Control</sub>* = Faktor externer präventiver Ursachen (hier nur die Kontrollbedingungen, d.h. keine anwesenden beobachteten präventiven Ursachen: „offene Türen“ vs. „keine Wände“); *C* = Status der gemeinsamen Ursache (zwei Stufen: nichts vs. „POR“); *N<sub>E+</sub>* = Anzahl der weiteren an „POR“ denkenden Aliens (drei Stufen: 0, 1 und 2).

\*\**p* < .01

Tabelle I-2

Varianzanalyse für Experiment 6 (Experimentalbedingungen)

Varianzquelle	<i>df</i>	<i>MSE</i>	<i>F</i>	<i>p</i>	$\eta^2$
Zwischen Probanden					
<i>EXT</i>	1	1.88	0.20	.657	.01
Fehler	30	9.33			
Innerhalb von Probanden (randomisierte Messwiederholung)					
<i>C</i>	1	641.67	77.09**	<.001	.72
<i>C x EXT</i>	1	18.13	2.18	.150	.07
Fehler	30	8.32			
<i>N<sub>E+</sub></i>	2	23.77	9.93**	<.001	.25
<i>N<sub>E+</sub> x EXT</i>	2	0.10	0.04	.960	.00
Fehler	60	2.39			
<i>C x N<sub>E+</sub></i>	2	23.95	11.90**	<.001	.28
<i>C x N<sub>E+</sub> x EXT</i>	2	2.66	1.32	.274	.04
Fehler	60	2.01			

*Anmerkungen.* *EXT* = Faktor externer präventiver Ursachen (hier nur die beiden relevanten Experimentalbedingungen: „Wände“ [präventive Ursache anwesend] vs. „offene Türen“ [präventive Ursache abwesend]); *C* = Status der gemeinsamen Ursache (zwei Stufen: nichts vs. „POR“); *N<sub>E+</sub>* = Anzahl der weiteren an „POR“ denkenden Aliens (drei Stufen: 0, 1 und 2).

\*\**p* < .01

Tabelle I-3

Konditionale Varianzanalyse für Experiment 6 (nur Experimentalbedingungen) für Faktor C=0 (nichts)

Varianzquelle	df	MSE	F	P	$\eta^2$
Zwischen Probanden					
EXT	1	4.17	0.40	.534	.01
Fehler	30	10.54			
Innerhalb von Probanden (randomisierte Messwiederholung)					
$N_{E+}$	2	0.09	0.10	.907	<.01
$N_{E+} \times EXT$	2	1.14	1.18	.313	.04
Fehler	60	0.96			

Anmerkungen. EXT = Faktor externer präventiver Ursachen (hier nur die beiden relevanten Experimentalbedingungen: „Wände“ [präventive Ursache anwesend] vs. „offene Türen“ [präventive Ursache abwesend]);  $N_{E+}$  = Anzahl der weiteren an „POR“ denkenden Aliens (drei Stufen: 0, 1 und 2).

Tabelle I-4

Konditionale Varianzanalyse für Experiment 6 (nur Experimentalbedingungen) für Faktor C=1 („POR“)

Varianzquelle	df	MSE	F	P	$\eta^2$
Zwischen Probanden					
EXT	1	15.84	2.23	.146	.07
Fehler	30	7.11			
Innerhalb von Probanden (randomisierte Messwiederholung)					
$N_{E+}$	1	47.63	13.82**	<.001	.32
$N_{E+} \times EXT$	2	1.63	0.47	.626	.02
Fehler	60	3.45			

Anmerkungen. EXT = Faktor externer präventiver Ursachen (hier nur die beiden relevanten Experimentalbedingungen: „Wände“ [präventive Ursache anwesend] vs. „offene Türen“ [präventive Ursache abwesend]);  $N_{E+}$  = Anzahl der weiteren an „POR“ denkenden Aliens (drei Stufen: 0, 1 und 2).

\*\*p < .01

## Anhang J – Varianzanalytische Ergebnisse – Experiment 7

Tabelle J-1

### Varianzanalyse für Experiment 7

Varianzquelle	<i>df</i>	<i>MSE</i>	<i>F</i>	<i>p</i>	$\eta^2$
Innerhalb von Probanden (randomisierte Messwiederholung)					
<i>EXT</i>	1	16.67	7.60*	.011	.25
Fehler	23	2.20			
-----					
$N_{E+}$	2	46.86	7.08**	.002	.24
Fehler	46	6.62			
-----					
$N_{E+} \times EXT$	2	4.36	3.42*	.041	.13
Fehler	46	1.27			

*Anmerkungen.* *EXT* = Faktor externer präventiver Ursachen („Wände“ [präventive Ursache anwesend] vs. „offene Türen“ [präventive Ursache abwesend]);  $N_{E+}$  = Anzahl der weiteren an „POR“ denkenden Aliens (drei Stufen: 0, 1 und 2).

\* $p < .05$ , \*\* $p < .01$

## Anhang K – Ausführliche Herleitung des erweiterten Modells

Es sei wieder  $C \in \{0,1\}$  eine binäre Ursache und  $E_1, \dots, E_n$  mit  $E_i \in \{0,1\} \forall i \in \{1, \dots, n\}$  binäre Effekte dieser Ursache (*Common-Cause-Struktur*),  $PN_j \in \{0,1\}$  unbeobachtete präventive Ursachen und  $w = (w_C, w_{PN}, b_C, b_{PN}, b_E)$  der zugehörige Parametervektor. – Für die genauere Definition und die *A-priori*-Verteilungen siehe Anhang D entsprechend. Sei des Weiteren  $F = (f_1, \dots, f_n)$  mit  $f_i \in \{A, B\} \forall i \in \{1, \dots, n\}$  in diesem Fall eine *Feature-Matrix* mit einem binären Merkmal pro Effekt<sup>68</sup>. Das Merkmal  $f_i$  ist dabei jeweils dem Effekt  $E_i$  zugeordnet; sei überdies  $\psi$  der Hyperparametervektor, der das kategoriale Wissen beschreibt (siehe zur Definition weiter unten), dieses sei unabhängig von den kausalen Parametern und betreffe daher nur den Kategorisierungsteil der Inferenz.

Seien Ursache  $C$  sowie die Effekte  $E_1, \dots, E_{n-1}$  und *alle* Merkmale in  $F$  beobachtet, der Zustand des Effekts  $E_n$  jedoch unbeobachtet (Zieleffekt). Dann gilt für die Wahrscheinlichkeit der Anwesenheit des Zieleffekts ( $E_n = 1$ ) gegeben der beobachteten Variablen:

$$\begin{aligned} P(E_n = 1 | C, E_1, \dots, E_{n-1}, F; \omega, \psi) \\ = \iint P(E_n = 1 | C, E_1, \dots, E_{n-1}, F; w, \psi) \cdot P(w | \omega) dw \end{aligned} \quad (\text{K-1})$$

unter der Annahme, dass<sup>69</sup>

$$P(w | C, E_1, \dots, E_{n-1}, F; \omega, \psi) = P(w | \omega) \quad (\text{K-2})$$

Sei überdies  $z$  eine beliebige Partitionierung der Effekte  $E_1, \dots, E_n$  in mindestens einen und höchstens  $n$  Cluster  $z_j$  und  $P(z | F; \psi)$  die (noch genauer zu definierende) Wahrscheinlichkeit dieser Partitionierung gegeben der Merkmale in  $F$ . Dann lässt sich die bedingte Wahrscheinlichkeit  $P(E_n = 1 | C, E_1, \dots, E_{n-1}, F; w, \psi)$  auch schreiben als gewichtete Summe über alle möglichen Partitionierungen:

$$\begin{aligned} P(E_n = 1 | C, E_1, \dots, E_{n-1}, F; w, \psi) \\ = \sum_z P(E_n = 1 | C, PN, E_1, \dots, E_{n-1}, z; w) \cdot P(z | F; \psi) \end{aligned} \quad (\text{K-3})$$

<sup>68</sup> Der Fall lässt sich einfach auf  $k$  Merkmale erweitern.

<sup>69</sup> Zur Begründung dieser Annahme siehe 4.2.

Innerhalb einer gegebenen Partitionierung  $z$  gebe es für jeden Cluster  $z_j$  eine gemeinsame Fehlerquelle  $PN_{z_j}$ , die dem Basismodell entsprechend präventiv auf die Ursache-Effekt-Relation wirkt. Seien nun  $E_{z_j}$  diejenigen Effekte, die zum  $j$ ten Cluster innerhalb der Partitionierung  $z$  gehören ( $E_n$ , der Zieleffekt, selbst nicht eingeschlossen) und  $z_n$  der Cluster der  $E_n$  enthält und damit  $E_{z_n}$  diejenigen Effekte, die sich im selben Cluster wie  $E_n$  befinden. Da  $E_n$  nur mit den Effekten  $E_{z_n}$  eine gemeinsame Fehlerquelle,  $PN_{z_n}$ , teilt, nicht aber mit den anderen Effekten der *Common-Cause*-Struktur, lässt sich die Zielinferenz gegeben einer Partitionierung  $z$  analog zur Herleitung in Anhang D vereinfacht schreiben als:

$$\begin{aligned} P(E_n = 1|C, E_1, \dots, E_{n-1}, z; w) \\ = \sum_{PN_{z_n}} P(E_n = 1|C, PN_{z_n}; w) \cdot P(PN_{z_n}|C, E_{z_n}; w) \end{aligned} \quad (\text{K-4})$$

Für die bedingte Wahrscheinlichkeit  $P(E_n = 1|C, PN_{z_j}; w)$  gelte die Wahrscheinlichkeitstabelle aus Anhang D entsprechend. Dann gilt analog der Herleitung des Basismodells mit  $m_{z_n}$  als der Anzahl der als anwesend beobachteten Effekte in Cluster  $z_n$  und  $n_{z_n}$  als der Anzahl der zum Cluster  $z_n$  gehörenden Effekte Gleichung D-12 entsprechend:

$$\begin{aligned} P(E_n = 1|C, E_1, \dots, E_{n-1}, z; w) \\ = \frac{\sum_{PN_{z_n}} P(E = 1|C, PN_{z_n}; w)^{m_{z_n}+1} \cdot P(E = 0|C, PN_{z_n}; w)^{n_{z_n}-m_{z_n}-1} \cdot P(PN_{z_n}|w)}{\sum_{PN_{z_n}} P(E = 1|C, PN_{z_n}; w)^{m_{z_n}} \cdot P(E = 0|C, PN_{z_n}; w)^{n_{z_n}-m_{z_n}-1} \cdot P(PN_{z_n}|w)} \end{aligned} \quad (\text{K-5})$$

Für die Wahrscheinlichkeit einer Partitionierung gegeben der Merkmale  $P(z|F)$  gilt nach Anwendung der Bayes-Regel:

$$P(z|F; \psi) \propto P(F|z; \psi) \cdot P(z) \quad (\text{K-6})$$

Die *A-priori*-Wahrscheinlichkeit einer Partitionierung  $P(z)$  lässt sich als *Chinese-Restaurant-Prozess* (CRP) beschreiben (siehe Pitman, 2002; Kemp, Goodman & Tenenbaum, 2007), der jeder Partitionierung  $z$  von  $n$  Objekten auf der Basis der Größe  $|z_j|$  der Cluster  $z_j$  in  $z$  eine Wahrscheinlichkeit dergestalt zuordnet, dass:

$$P(z) = \frac{1}{n!} \prod_{z_j \in z} (|z_j| - 1)! \quad (\text{K-7})$$

Das Likelihood einer Merkmalskonfiguration  $F$  gegeben einer bestimmten Partitionierung  $z$  lässt sich als Beta-Binomial-Modell beschreiben. Seien dazu  $F_{z_j}$  die Merkmale  $f_i$ , die zu den in Cluster  $z_j$  befindlichen Merkmalen gehören und darüber hinaus jedem Cluster  $z_j$  eine Wahrscheinlichkeit  $\theta_{z_j}$  zugeordnet, so dass

$$\theta_{z_j} = P(f_i = A) \forall f_i \in F_{z_j} \quad (\text{K-8})$$

Dann folgt:

$$P(F|z; \psi) = \prod_{z_j} P(F_{z_j}|\psi) = \prod_{z_j} \int P(F_{z_j}|\theta_{z_j}) \cdot P(\theta_{z_j}|\psi) d\theta_{z_j} \quad (\text{K-9})$$

mit

$$P(F_{z_j}|\theta_{z_j}) = \binom{|z_j|}{\#(F_{z_j} = A)} (\theta_{z_j})^{\#(F_{z_j} = A)} \cdot (1 - \theta_{z_j})^{\#(F_{z_j} = B)} \quad (\text{K-10})$$

als Binomial-Modell und

$$P(\theta_{z_j}|\psi) = \frac{\Gamma(\alpha_\theta + \beta_\theta)}{\Gamma(\alpha_\theta)\Gamma(\beta_\theta)} (\theta_{z_j})^{\alpha_\theta - 1} \cdot (1 - \theta_{z_j})^{\beta_\theta - 1} \sim \text{Beta}(\alpha_\theta, \beta_\theta) \quad (\text{K-11})$$

als Beta-Prior mit den Hyperparametern  $\alpha_\theta$  und  $\beta_\theta$  und damit  $\psi = (\alpha_\theta, \beta_\theta)$ .

Da die Beta-Verteilung konjugiert ist zum Binomial-Modell (siehe Gelman et al., 2004), lässt sich Gleichung K-9 analytisch lösen:

$$\begin{aligned} P(F|z; \psi) &= \prod_{z_j} P(F_{z_j}|\psi) \\ &= \prod_{z_j} \int \binom{|z_j|}{\#(F_{z_j} = A)} (\theta_{z_j})^{\#(F_{z_j} = A)} \cdot (1 - \theta_{z_j})^{\#(F_{z_j} = B)} \cdot \frac{\Gamma(\alpha_\theta + \beta_\theta)}{\Gamma(\alpha_\theta)\Gamma(\beta_\theta)} (\theta_{z_j})^{\alpha_\theta - 1} \cdot (1 - \theta_{z_j})^{\beta_\theta - 1} d\theta_{z_j} \quad (\text{K-12}) \end{aligned}$$

$$= \prod_{z_j} \int \binom{|z_j|}{\#(F_{z_j} = A)} \cdot \frac{\Gamma(\alpha_\theta + \beta_\theta)}{\Gamma(\alpha_\theta)\Gamma(\beta_\theta)} \cdot (\theta_{z_j})^{\#(F_{z_j} = A) + \alpha_\theta - 1} \cdot (1 - \theta_{z_j})^{\#(F_{z_j} = B) + \beta_\theta - 1} d\theta_{z_j} \quad (\text{K-13})$$

$$= \prod_{z_j} \binom{|z_j|}{\#(F_{z_j} = A)} \cdot \frac{\Gamma(\alpha_\theta + \beta_\theta)}{\Gamma(\alpha_\theta)\Gamma(\beta_\theta)} \cdot \frac{\Gamma(\alpha_\theta + \#(F_{z_j} = A))\Gamma(\beta_\theta + \#(F_{z_j} = B))}{\Gamma(\alpha_\theta + \beta_\theta + |z_j|)} \quad (\text{K-14})$$

$$= \prod_{z_j} \binom{|z_j|}{\#(F_{z_j} = A)} \cdot \frac{\Gamma(\alpha_\theta + \beta_\theta)}{\Gamma(\alpha_\theta)\Gamma(\beta_\theta)} \cdot \frac{\Gamma(\alpha_\theta + \#(F_{z_j} = A))\Gamma(\beta_\theta + \#(F_{z_j} = B))}{\Gamma(\alpha_\theta + \beta_\theta + |z_j|)} \quad (\text{K-15})$$

Zusammen mit dem CRP-Prior ergibt sich also für die Posterior-Wahrscheinlichkeit einer Partitionierung:

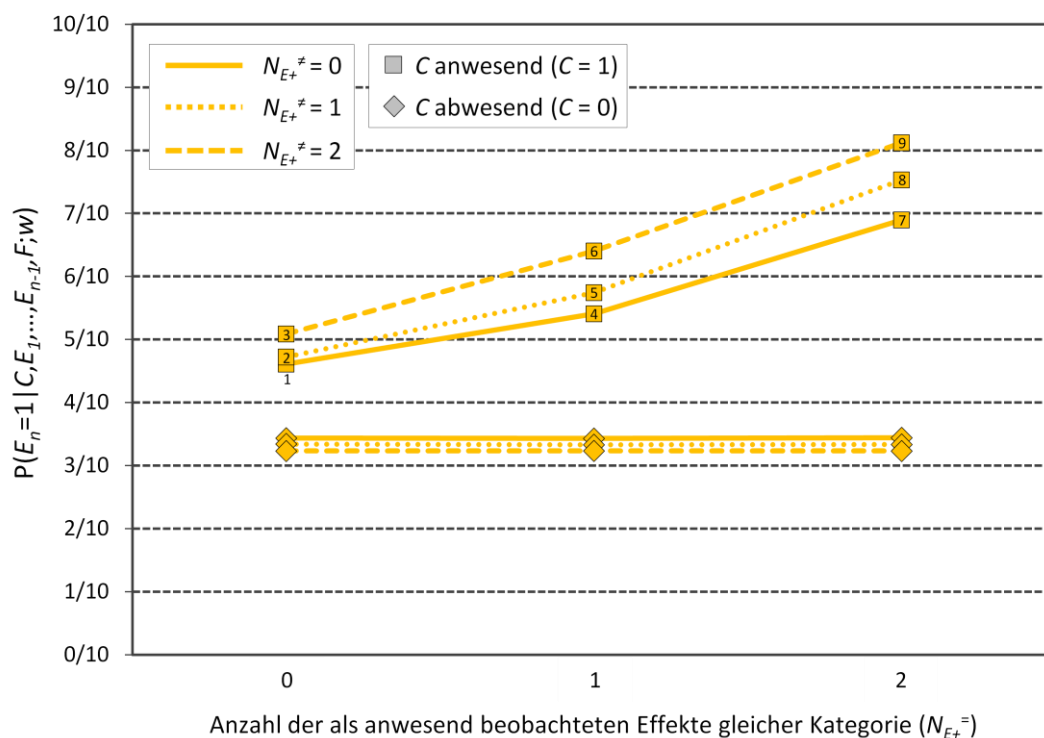
$$P(z|F; \psi) \propto \frac{1}{n!} \prod_{z_j} \binom{|z_j|}{\#(F_{z_j} = A)} \cdot \frac{\Gamma(\alpha_\theta + \beta_\theta)}{\Gamma(\alpha_\theta)\Gamma(\beta_\theta)} \cdot \frac{\Gamma(\alpha_\theta + \#(F_{z_j} = A))\Gamma(\beta_\theta + \#(F_{z_j} = B))}{\Gamma(\alpha_\theta + \beta_\theta + |z_j|)} \cdot \Gamma(|z_j|) \quad (\text{K-16})$$

Die Zielinferenz selbst (Gleichung K-1) kann dabei analytisch nicht gelöst, aber analog zu dem Vorgehen in Anhang D durch Monte-Carlo-Simulation approximiert werden. Da die Inferenz selbst auch eine gewichtete Summe ist, kann diese vereinfacht simuliert werden, in dem für jede Partitionierung  $z$  ein  $M_z$  gewählt wird, so dass  $M = \sum_z M_z$  und  $\frac{M_z}{M} = P(z|F)$ . Die Approximation erfolgt dann gemäß Anhang D für Gleichung K-5 entsprechend.



## Anhang L – Detaillierte Vorhersage des erweiterten Modells

In Abbildung L-1 und Abbildung L-2 sind in Ergänzung zu Abbildung 38 in Abschnitt 6.2.4 die detaillierten Vorhersagen des erweiterten Modells dargestellt. Es handelt sich bei beiden Abbildungen um die Vorhersage der gleichen Datenpunkte, zum einen in Abhängigkeit der Anzahl der als anwesend beobachteten Effekte gleicher Kategorie gruppiert, nach der Anzahl der anwesend beobachteten Effekte der anderen Kategorie (Abbildung L-1) und zum anderen genau anders herum (Abbildung L-2). Zur Verdeutlichung, dass es sich um die gleichen Datenpunkte handelt, sind die Datenpunkte für den Fall der Anwesenheit der Ursache (die jeweils oberen drei Linien) in beiden Abbildungen von 1 bis 9 nummeriert.



*Abbildung L-1.* Vorhersage des erweiterten Modells auf der Basis einer Monte-Carlo-Simulation mit 100 000 Durchgängen (siehe weiteres in Abschnitt 6.2.4 und Abbildung 38) in Abhängigkeit von der Anzahl der als anwesend beobachteten Effekte der gleichen Kategorie gruppiert nach der Anzahl der als anwesend beobachteten Effekte der anderen Kategorie sowie dem Zustand der Ursache  $C$  (anwesend vs. abwesend). Der Durchschnitt der Linien je Zustand der Ursache entspricht den *in Abbildung 38 abgetragenen gestrichelten, gelben Linien*; also für die Anwesenheit der Ursache der Mittelung über die Punkte:  $\overline{(1,2,3)}-\overline{(4,5,6)}-\overline{(7,8,9)}$ .

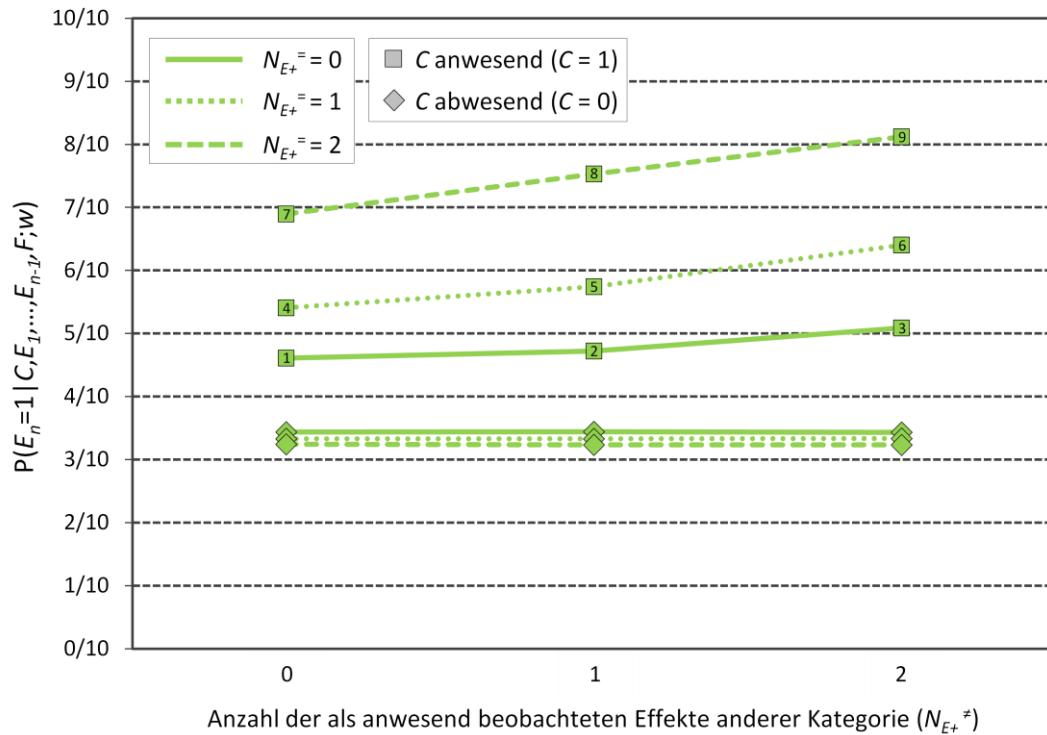
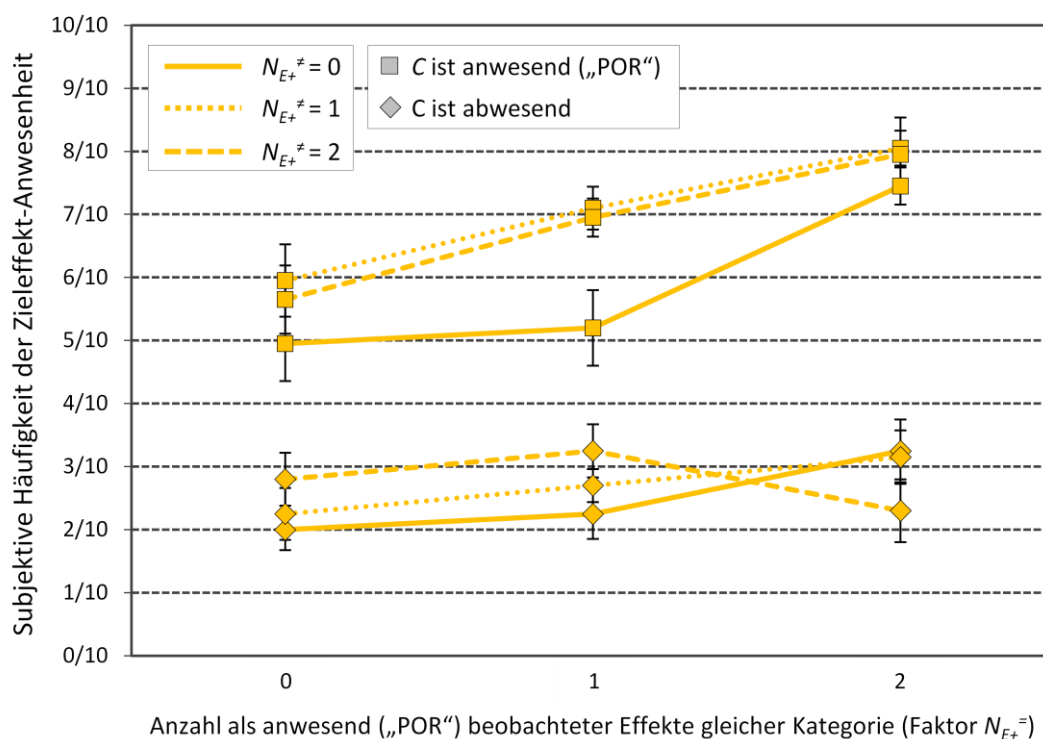


Abbildung L-2. Vorhersage des erweiterten Modells auf der Basis einer Monte-Carlo-Simulation mit 100 000 Durchgängen (siehe weiteres in Abschnitt 6.2.4 und Abbildung 38) in Abhängigkeit von der Anzahl der als anwesend beobachteten Effekte der anderen Kategorie gruppiert nach der Anzahl der als anwesend beobachteten Effekte der gleichen Kategorie sowie dem Zustand der Ursache  $C$  (anwesend vs. abwesend). Der Durchschnitt der Linien je Zustand der Ursache entspricht den in Abbildung 38 abgetragenen durchgezogenen, grünen Linien; also für die Anwesenheit der Ursache der Mittelung über die Punkte:  $\overline{(1,4,7)}$ – $\overline{(2,5,8)}$ – $\overline{(3,6,9)}$ .

## Anhang M – Detaillierte Ergebnisse von Experiment 8

In Abbildung M-1 und Abbildung M-2 sind in Ergänzung zu Abbildung 40 in Abschnitt 6.3.2 die detaillierten Ergebnisse von Experiment 8 dargestellt; weitere Erläuterungen siehe Abschnitt 6.3.2 und bezüglich der Aggregation Anhang L. Man beachte: Beide Abbildungen stellen die gleichen Datenpunkte nur jeweils in anderer Gruppierung dar.



*Abbildung M-1.* Durchschnittliche Ratings bezüglich der Häufigkeit der Anwesenheit des Zieleffekts in zehn fiktiven Situationen gruppiert nach dem Status der gemeinsamen Ursache  $C$  und der Anzahl der als anwesend beobachteten weiteren Effekte der anderen Kategorie (Faktor  $N_{E+}^{\neq}$ ) in Abhängigkeit von der Anzahl der als anwesend beobachteten Effekte der gleichen Kategorie (Faktor  $N_{E+}^{\neq}$ ) in Experiment 8 ( $N = 60$ ; balanciertes partiell-konfundiertes randomisiertes Block-Design). Die Fehlerbalken stellen den Standardfehler des Mittelwerts dar.

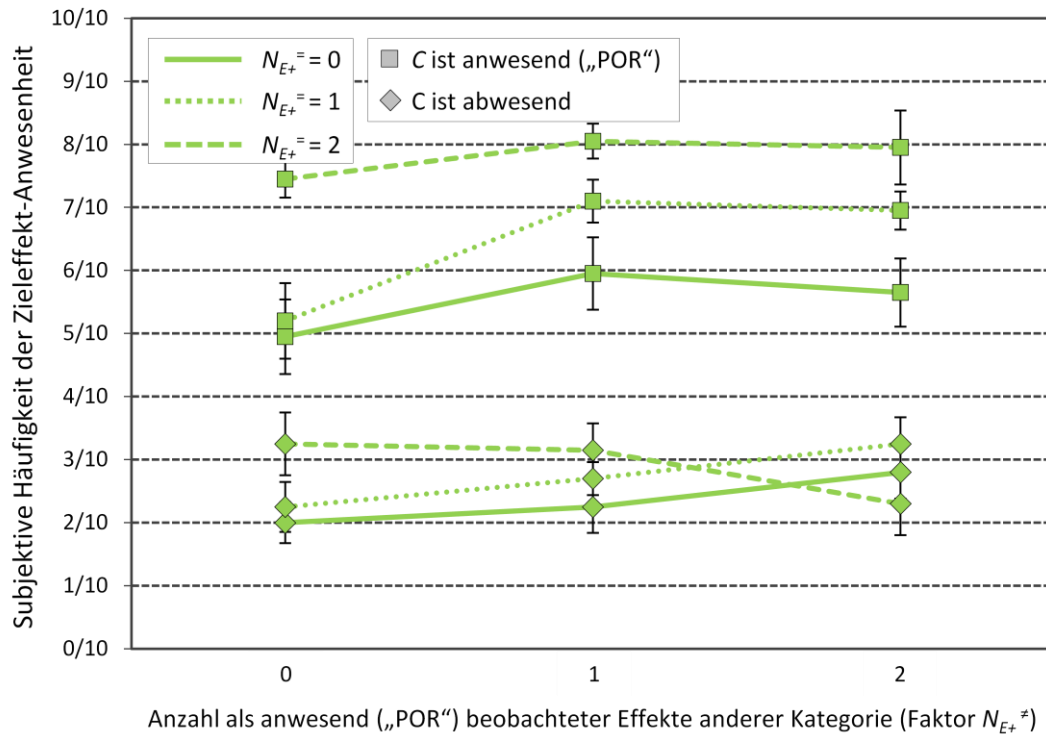


Abbildung M-2. Durchschnittliche Ratings bezüglich der Häufigkeit der Anwesenheit des Zieleffekts in zehn fiktiven Situationen gruppiert nach dem Status der gemeinsamen Ursache C und der Anzahl der als anwesend beobachteten weiteren Effekte der gleichen Kategorie (Faktor  $N_{E+}^-$ ) in Abhängigkeit von der Anzahl der als anwesend beobachteten Effekte der anderen Kategorie (Faktor  $N_{E+}^{\neq}$ ) in Experiment 8 (N = 60; balanciertes partiell-konfundiertes randomisiertes Block-Design). Die Fehlerbalken stellen den Standardfehler des Mittelwerts dar.

## Anhang N – Varianzanalytische Ergebnisse für Experiment 8

Tabelle N-1

Konditionale Varianzanalyse für Experiment 8 für Faktor C=0

Varianzquelle	df	MSE	F	P	$\eta^2$
Innerhalb von Probanden (balanciertes, partiell-konfundiertes randomisiertes Block-Design)					
$N_{E+}^-$	2	4.77	1.91	.152	.03
$N_{E+}^\neq$	2	1.27	0.51	.602	<.01
$N_{E+}^- \times N_{E+}^\neq$	4	2.50	1.00	.410	.04
Fehler	112	2.49			

Anmerkungen.  $N_{E+}^-$  = Anzahl der weiteren an „POR“ denkenden Aliens der gleichen Kategorie wie der Zieleffekt (drei Stufen: 0, 1 und 2);  $N_{E+}^\neq$  = Anzahl der weiteren an „POR“ denkenden Aliens der anderen Kategorie als der Zieleffekt (drei Stufen: 0, 1 und 2).

Tabelle N-2

Konditionale Varianzanalyse für Experiment 8 für Faktor C=1

Varianzquelle	df	MSE	F	P	$\eta^2$
Innerhalb von Probanden (balanciertes, partiell-konfundiertes randomisiertes Block-Design)					
$N_{E+}^-$	2	80.60	22.53**	<.001	.26
$N_{E+}^\neq$	2	23.62	6.60**	.002	.08
$N_{E+}^- \times N_{E+}^\neq$	4	5.16	1.44	.272	.01
Fehler	112	3.58			

Anmerkungen.  $N_{E+}^-$  = Anzahl der weiteren an „POR“ denkenden Aliens der gleichen Kategorie wie der Zieleffekt (drei Stufen: 0, 1 und 2);  $N_{E+}^\neq$  = Anzahl der weiteren an „POR“ denkenden Aliens der anderen Kategorie als der Zieleffekt (drei Stufen: 0, 1 und 2).

\*\*p < .01

## Lebenslauf

- Name: Ralf Mayrhofer
- 22.04.1977 geboren in Magdeburg
- 1983–1991 Tamara-Bunke-Schule Magdeburg [mit Unterbrechung  
 – 1985/86: Botschaftsschule in Aden (Jemen)  
 – 1988/89: Botschaftsschule in Addis Abeba (Äthiopien)]
- 1991–1995 Wilhelm-Raabe-Gymnasium Magdeburg, Abitur am 18.06.1995
- 1995–1996 Grundwehrdienst (Hildesheim, Achim)
- 1996–2004 Psychologie-Studium, Georg-August-Universität Göttingen;  
 Diplom im März 2004
- 1997–2005 Wirtschaftsinformatik-Studium, Georg-August-Universität  
 Göttingen; Diplom im Januar 2005
- 2001–2005 Jura-Studium, Georg-August-Universität Göttingen;  
 Zwischenprüfung im März 2005
- 2004–2007 Volkswirtschaftslehre-Studium, Georg-August-Universität  
 Göttingen, Diplom im Juli 2007
- seit 10/2004 PhD-Studiengang „Angewandte Statistik und Empirische Metho-  
 den“, Zentrum für Statistik, Georg-August-Universität Göttingen
- seit 11/2005 Wissenschaftlicher Mitarbeiter am Institut für Psychologie der  
 Universität Göttingen, Abteilung für Kognitionswissenschaft und  
 Entscheidungspsychologie (DFG-Projekt „Kausale Bayes-Netze als  
 psychologische Theorien kausalen Denkens“)
- Tätigkeiten in der akademischen und studentischen Selbstverwaltung der  
 Georg-August-Universität Göttingen (Auszug):
- 2002–2007 Mitglied des Senats
- 2003–2007 Mitglied des Studierendenparlaments
- 2005–2007 Präsident des Studierendenparlaments
- 2005–2007 Mitglied im GAUSS-Vorstand (Georg August University School of  
 Science)
- 2006–2007 Vorsitzender der erweiterten zentralen Kommission für Lehre  
 und Studium (zKLS-plus) [Studienbeitragsverwendung]
- 2006–2008 Vorsitzender des studentischen Wahlausschusses