



GEORG-AUGUST-UNIVERSITÄT  
GÖTTINGEN

**Gruppenleistung und Gruppenlernen  
bei der Steuerung dynamischer Systeme**

**Dissertation**

Zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades

„Doctor rerum naturalium“

der Georg-August-Universität Göttingen

im Promotionsprogramm Biologie

der Georg-August University School of Science (GAUSS)

vorgelegt von

**Sylvana Drewes**

aus Bad Gandersheim

Göttingen, 2013

### Betreuungsausschuss

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

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

### Mitglieder der Prüfungskommission

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

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

Weitere Mitglieder der Prüfungskommission:

Dr. Bernhard Fink, Heisenberg-Stipendiat, Courant Forschungszentrum Evolution des Sozialverhaltens

Prof. Dr. Andreas Glöckner, Abteilung Psychologische Diagnostik, Urteilen und Entscheiden, Georg-Elias-Müller-Institut für Psychologie

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

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

Tag der mündlichen Prüfung: 28.05.2013

## Danksagung

Mit der Fertigstellung dieser Dissertation endet eine Lern- und Arbeitsphase, in der ich durch Höhen und Tiefen gegangen bin, die die empirische Forschung mit sich bringen kann. Auf die von Vorfreude geprägte Einarbeitung in das Themengebiet „Gruppenleistung und Gruppenlernen“, die Eingrenzung der Fragestellung und die Festlegung der Methodik folgten bald erste Schwierigkeiten in Bezug auf die technische Umsetzbarkeit der in dieser Arbeit erstmals untersuchten computervernetzten Steuerung dynamischer Systeme durch Nominalgruppen. Die drei durchgeführten Gruppenexperimente waren durch eine aufwendige Planungsphase und eine zeitintensive Erhebungsphase gekennzeichnet, in denen Frustrationstoleranz und Geduld erforderlich waren. Während der Datenaufbereitung und -auswertung sowie der Ergebnisinterpretation sah ich mich (ähnlich wie die Probanden) mit Herausforderungen konfrontiert, die aus der Dynamik und Komplexität des gewählten Aufgabentyps der Systemsteuerung resultierten. Der Abschluss dieser Dissertation wäre ohne die unterstützende Begleitung einer Vielzahl von Menschen nicht möglich gewesen. Daher möchte ich ihnen meinen tiefen Dank aussprechen.

Allen voran danke ich meinem Doktorvater Stefan Schulz-Hardt, der mir ermöglicht hat, dieses aufwendige Projekt in der Abteilung Wirtschafts- und Sozialpsychologie zu realisieren. Er hat mich über alle Phasen der Dissertation hinweg mit Geduld und konstruktivem, fachkompetent-analytischen Feedback begleitet. Seinen präzisen wissenschaftlichen Arbeitsstil habe ich stets bewundert.

Mein Dank gilt auch Margarete Boos für die Bereitschaft, Zweitgutachterin meiner Arbeit zu sein sowie für ihre hilfreichen Anregungen im Entstehungsprozess der Arbeit. Bei Bernhard Fink, Andreas Glöckner, Uta Lass und Hannes Rakoczy möchte ich mich für die Mitwirkung in der Prüfungskommission für meine Dissertation bedanken. Bedanken möchte ich mich darüber hinaus bei Thomas Schultze, Frank Vogelgesang und Andreas Mojzisch, deren kollegialer Rat im Rahmen von Forschungskolloquien und persönlichen Gesprächen wichtige Impulse für diese Dissertation geliefert hat.

Ohne die Unterstützung von Hilfskräften wäre die Durchführung der Experimente, die die Grundlage für diese Arbeit bilden, nicht möglich gewesen. Insbesondere bin ich Alexander Schlemmer zu Dank verpflichtet, der das computerbasierte Framework sowie die spezifische Software für die vernetzten, dynamischen Systemsteuerungsaufgaben entwickelte. Ebenfalls danke ich der Diplomandin Anthea van Rooyen sowie den

Hilfskräften Stefanie Janssen, Liane Saucke, Franziska Lammers, Alexander Stern, Christoph Kwiatkowski, Anna Maria Warzecha und Sophie Kroh, die an der Datenerhebung bzw. der nachträglichen Kodierung beteiligt waren.

Mein Dank gilt der Konrad-Adenauer-Stiftung. Das Promotionsstipendium hat mir sowohl durch die finanzielle Förderung Freiräume eröffnet als auch im Rahmen der ideellen Förderung vielfältige interessante Erfahrungen und persönliche Begegnungen ermöglicht.

Ein großer Dank geht auch an Janina Braniella, Saskia Geisler und Jeremy Prasetyo, die mich insbesondere während der schwierigen Phasen des Promotionsvorhabens aufgemuntert und motiviert haben. Schließlich möchte ich meinen Eltern danken, die mich immer bedingungslos unterstützt haben. Ihnen ist diese Arbeit gewidmet.

*Göttingen, im Mai 2013*

*Sylvana Drewes*



# Inhalt

<b>1. Einleitung .....</b>	<b>1</b>
<b>2. Theoretischer und empirischer Hintergrund .....</b>	<b>4</b>
2.1 Begriffsklärung und aufgabenklassifikatorische Einordnung von dynamischen Systemen und Systemsteuerungsaufgaben.....	4
2.2 Forschungsstand zur Steuerung dynamischer Systeme .....	8
2.2.1 Systemsteuerungsleistung und systembezogener Wissenserwerb von Individuen .....	9
2.2.2 Systemsteuerungsleistung und systembezogener Wissenserwerb von Gruppen .....	13
2.3 Gruppenleistung und Gruppenlernen in der sozialpsychologischen Forschung ....	20
2.3.1 Bestimmung von Prozessgewinnen bei quantitativen Steuerungsurteilen...22	
2.3.2 Differenzierung von Gruppenlernprozessen bei wiederholter kollektiver Aufgabenbearbeitung .....	24
<b>3. Zielsetzung und Ableitung experimentübergreifender Hypothesen .....</b>	<b>26</b>
<b>4. Experiment 1 .....</b>	<b>31</b>
4.1 Zielsetzung und Hypothesen .....	31
4.2 Stichprobe und Design .....	31
4.3 Methode .....	32
4.3.1 Steuerungsaufgabe und Leistungsmaß .....	32
4.3.2 Experimenteller Ablauf.....	41
4.3.3 Erfassung des Wissenserwerbs.....	43
4.4 Ergebnisse und Diskussion .....	46
4.4.1 Gruppenleistung .....	46
4.4.2 Lerneffekte auf Gruppenebene .....	52
4.4.3 Lerneffekte auf individueller Ebene.....	56
4.4.4 Wissenserwerb bei der Systemsteuerung.....	59
4.4.5 Explorative Analysen: Erkennen von steuerungsbezogener Expertise und subjektive Gewichtung der Individualurteile im Realgruppenkontext ....	64
4.4.6 Zusammenfassung und Ausblick auf Experiment 2.....	71
<b>5. Experiment 2 .....</b>	<b>73</b>
5.1 Zielsetzung und Hypothesen .....	73
5.2 Stichprobe und Design .....	74

5.3	Methode .....	75
5.3.1	Steuerungsaufgabe und Leistungsmaß .....	75
5.3.2	Experimenteller Ablauf.....	82
5.3.3	Erfassung des Wissenserwerbs.....	83
5.4	Ergebnisse und Diskussion .....	87
5.4.1	Gruppenleistung .....	89
5.4.2	Lerneffekte auf Gruppenebene .....	93
5.4.3	Lerneffekte auf individueller Ebene.....	96
5.4.4	Wissenserwerb bei der Systemsteuerung.....	99
5.4.5	Explorative Analysen: Erkennen von steuerungsbezogener Expertise und subjektive Gewichtung im Realgruppenkontext.....	102
5.4.6	Zusammenfassung .....	106
<b>6.</b>	<b>Experiment 3 .....</b>	<b>109</b>
6.1	Zielsetzung und Hypothesen .....	109
6.2	Stichprobe und Design .....	111
6.3	Methode .....	112
6.3.1	Steuerungsaufgabe und Leistungsmaß .....	112
6.3.2	Experimenteller Ablauf.....	112
6.3.3	Erfassung des Wissenserwerbs.....	114
6.4	Ergebnisse und Diskussion .....	114
6.4.1	Gruppenleistung .....	115
6.4.2	Lerneffekte auf Gruppenebene .....	117
6.4.3	Lerneffekte auf individueller Ebene.....	121
6.4.4	G-I Transfer im Realgruppenkontext.....	125
6.4.5	Wissenserwerb bei der Systemsteuerung.....	129
6.4.6	Zusammenfassung.....	134
6.5	Metaanalyse zu Prozessgewinnen .....	136
<b>7.</b>	<b>Allgemeine Diskussion.....</b>	<b>138</b>
7.1	Gruppenleistung und Gruppenkoordination bei der Steuerung dynamischer Systeme .....	139
7.2	Leistungs- und wissensbezogener Group-to-Individual Transfer bei der Steuerung dynamischer Systeme .....	141
7.3	Beschränkungen und Implikationen für die Praxis und weitere Forschung .....	143
7.4	Abschließende Bemerkung.....	153
	<b>Literatur.....</b>	<b>155</b>

<b>Tabellenverzeichnis .....</b>	<b>168</b>
<b>Abbildungsverzeichnis .....</b>	<b>170</b>
<b>Anhang .....</b>	<b>174</b>



## 1. Einleitung

Am 26.04.1986 ereignete sich eine Katastrophe, an deren Folgen nach Expertenschätzungen zwischen 8930 und 1.44 Millionen Menschen starben (Krassuski, 2011). Der Reaktor 4 des Kernkraftwerkes im ukrainischen Tschernobyl explodierte. Bis heute erhöht die dabei freigesetzte radioaktive Strahlung das Krebsrisiko unzähliger Menschen. Diese Katastrophe war nicht auf technisches Versagen oder eine spezifische Fehlentscheidung des Sicherheitspersonals zurückzuführen. Stattdessen führten viele Fehlerurteile von Operateuren und Ingenieuren, die wegen eines Sicherheitstests über mehrere Stunden manuell in die Steuerung des Reaktors und des zugehörigen Kühlsystems eingriffen, zu einer destabilisierenden Übersteuerung des Systems und schließlich zur Reaktorexlosion (Hofinger, Rek, & Strohschneider, 2006). Der Atomreaktor von Tschernobyl stellte ein komplexes dynamisches System dar, in dem Steuerungseingriffe zahlreiche Systemparameter beeinflussten, und parallel unbeabsichtigte Nebeneffekte und zeitverzögerte negative Folgen hervorriefen. Die nachträgliche Analyse des Steuerungsprozesses ergab nach Hofinger et al. (2006), dass die Explosion zu mehreren Zeitpunkten aufzuhalten gewesen wäre, wenn die steuernden Individuen die Komplexität und Dynamik des Systems angemessen berücksichtigt hätten.

Die Tatsache, dass mangelnde menschliche Systemsteuerungsleistungen maßgeblich an der Nuklearkatastrophe beteiligt waren, ist besonders besorgniserregend, da die Steuerung dynamischer Systeme in vielen weiteren sicherheitstechnisch, ökologisch und wirtschaftlich relevanten Bereichen an der Tagesordnung ist. Menschen greifen dabei gezielt in komplexe Variablengefüge ein, um bestimmte günstige oder optimale Systemzustände herbeizuführen oder beizubehalten. Die intensivierete Industrialisierung mit dem Ziel, Wirtschaftswachstum und Bevölkerungswohlstand zu erhöhen, ist ein weiteres globales Beispiel für negative Folgen von Systemeingriffen in Form des Treibhauseffektes und der globalen Erwärmung. In Wirtschaftssystemen ist das Systemsteuerungsziel meist die Maximierung der Profitabilität. Schlechte Steuerungsleistungen im beruflichen Kontext haben häufig weitreichende Konsequenzen. Managementurteile zur Anpassung der Produktion an sich verändernde Marktbedingungen können bei hoher Urteilsqualität zu Unternehmenserfolg und Sicherung von Arbeitsplätzen führen - oder aber bei suboptimalen Urteilen Insolvenz und Arbeitsplatzabbau nach sich ziehen. Entsprechend wichtig sind angemessene Steuerungsurteile, die zu gewünschten positiven Systemzuständen führen und keine ungewollten Nebeneffekte hervorrufen.

Neben den Wirtschaftswissenschaften hat sich insbesondere die Psychologie mit der Frage beschäftigt, wie Verantwortliche bei der Steuerung vorgehen, welche Steuerungsleistungen sie gemessen an den vorgegebenen Zielen oder normativen Vergleichsmaßstäben erbringen, welche Urteilsfehler ihnen dabei unterlaufen und wie der systembezogene Wissenserwerb erfolgt. Der größte Teil der empirischen Forschung bezieht sich dabei auf den individuellen Umgang mit dynamischen Systemen und zeigt Defizite in der Steuerungsleistung auf (z.B. Brehmer, 2005; Funke, 1992). Individuen haben unter anderem Schwierigkeiten non-lineare Beziehungen zwischen Systemvariablen zu erkennen und unerwünschte Nebeneffekte sowie langfristige Folgen ihrer Systemeingriffe zu antizipieren, die zur Instabilität der Systeme oder unnötig hohen Kosten führen (Paich & Sterman, 1993).

Zur Bearbeitung von Systemsteuerungsaufgaben in der beruflichen Praxis werden häufig Managementteams oder Gruppen von Sicherheitsexperten eingesetzt. Ein Grund für den Einsatz von Gruppen ist die implizite Annahme, dass durch die aufgabenbezogene Interaktion in der Gruppe Synergieeffekte auftreten und die Gruppe eher Fehler vermeidet und zielführendere, besser an die sich dynamisch verändernde Situation angepasste Urteile trifft als Individuen (vgl. Burke, Stagl, Salas, Pierce, & Kendall, 2006; Srivastava & Bhattacharya, 2007). Die bisherige Forschung hat über unterschiedliche Gruppenaufgaben hinweg häufiger Prozessverluste in Gruppen, also ein schlechteres Abschneiden der Gruppe als auf Grund der individuellen Leistungen zu erwarten wäre, nachgewiesen als Prozessgewinne (z.B. Kerr & Bruun, 1983; Ingham, Levinger, Graves und Peckham, 1974). Allerdings finden sich bei der Bearbeitung statischer logischer Problemlöseaufgaben wie Regelinduktionsaufgaben Belege dafür, dass Gruppenurteile zu besseren Ergebnissen führen als der Mittelwert der individuellen Urteile der Gruppenmitglieder (Laughlin, VanderStoep, & Hollingshead, 1991). Bisher wurde noch nicht systematisch untersucht, ob soziale Interaktion auch bei dynamischen Systemsteuerungsaufgaben zu Leistungsvorteilen gegenüber der individuellen Steuerung führt. An dieser Forschungslücke setzt diese Arbeit an.

In meiner Dissertation untersuche ich in einer Serie von drei Experimenten erstmals methodisch sauber, ob Gruppen bei der Steuerung dynamischer Systeme interaktionsbedingte Leistungsvorteile gegenüber der Leistung einer vergleichbaren Anzahl an Individuen, so genannter Nominalgruppen, realisieren können. Darüber hinaus werde ich klären, ob die erzielte Gruppenleistung vollständig auf dem Vorteil der statistischen Aggregation der individuellen Urteile basiert - ein Phänomen das als *assembly effect bonus* (Collins & Guetzkow, 1964) oder *wisdom of the crowds* (Surowiecki, 2004) bezeichnet wird

- oder ob die Gruppeninteraktion darüber hinaus zu Synergieeffekten führt. In weiteren Analysen vergleiche ich die Gruppenleistung mit etablierten normativen Modellen und exploriere, ob Gruppen das Leistungsniveau des besten Gruppenmitglieds oder der situationsspezifisch besten individuellen Steuerungsurteile erreichen. Darüber hinaus wird auf Basis des dynamischen Modells der Gruppenleistung von Brodbeck und Greitemeyer (2000a) empirisch geprüft, welchen Anteil im Gruppenkontext auftretende Lernprozesse an der gruppenurteilsbasierten Steuerungsleistung haben. Während Experiment 1 eine Kühlsystemsteuerungsaufgabe in einem Atomkraftwerkszenario beinhaltet, wird in Experiment 2 die Generalisierbarkeit der Befunde auf ein dynamisches Wirtschaftssystem mit größerer Komplexität untersucht. Experiment 3 ermöglicht mit einem erweiterten Design des Atomkraftwerkszenarios eindeutig stabilen interaktionsbedingten Lerntransfer auf individueller Ebene nachzuweisen. In allen Experimenten wird darüber hinaus untersucht, ob sich die Gruppeninteraktion förderlich auf den qualitativen und quantitativen Erwerb von Systemwissen auswirkt.

Die Arbeit ist folgendermaßen gegliedert: In Kapitel 2 werden die allgemeinen Charakteristika dynamischer Systeme und Systemsteuerungsaufgaben beschrieben. Bei der anschließenden Darstellung des Forschungsstands zur Steuerungsleistung von Individuen und Gruppen werden methodische Mängel der Studien und daraus resultierende Forschungslücken in Bezug auf die Gruppenleistung und in der Gruppe auftretende Lernprozesse herausgestellt. Daran anknüpfend beschreibe ich eine sozialpsychologische Herangehensweise an die Untersuchung von Gruppenleistung und Lernprozessen und fokussiere dabei die Differenzierung von möglichen Lernprozessen im Gruppenkontext. In Kapitel 3 werden die Forschungsziele und die experimentübergreifenden Hypothesen formuliert. Die Kapitel 4 bis 6 beschreiben Zielsetzung, Methodik, Ergebnisse sowie die experimentspezifische Ergebnisdiskussion der drei durchgeführten Experimente. Kapitel 6 schließt mit einer Integration der Ergebnisse zu Prozessgewinnen bei der Systemsteuerung von Gruppen aus allen durchgeführten Experimenten in einer Metaanalyse ab. Kapitel 7 beinhaltet die allgemeine Diskussion der Ergebnisse in Relation zu bisheriger Forschung und zeigt die Grenzen dieser Arbeit auf. Schließlich werden Implikationen für die weitere Forschung und den Einsatz von Gruppen und möglichen gruppenunterstützenden Interventionen bei der hochgradig relevanten Systemsteuerung in der Praxis abgeleitet.

## 2. Theoretischer und empirischer Hintergrund

### 2.1 Begriffsklärung und aufgabenklassifikatorische Einordnung von dynamischen Systemen und Systemsteuerungsaufgaben

Die wissenschaftliche Untersuchung menschlichen Verhaltens bei der Steuerung dynamischer Systeme fand in den 1970er Jahren Einzug in die Psychologie. Die zentrale Forschungsmethodik stellen bis heute (meist vereinfachte) Computersimulationen wirtschaftlicher, ökologischer, biologischer oder technischer Systeme dar, mit denen Steuerungsverhalten und Leistung von Personen in laborexperimentellen Untersuchungen erfasst werden. Zunächst beschäftigten sich vorrangig Alltagspsychologen im Rahmen der Denkpsychologie mit menschlicher Leistung im Kontext dynamischer Systeme (z.B. Dörner, Kreuzig, Reither, & Stäudel, 1983; Funke, 1985; Kluwe, 1995).<sup>1</sup> Dynamische Systemsteuerungsaufgaben werden in der Forschungstradition der Denkpsychologie als *komplexe Probleme* bezeichnet. Funke (1992) erläutert diesen Begriff folgendermaßen:

Das Adjektiv „komplex“ charakterisiert somit eine wesentliche Eigenart des verwendeten Stimulusmaterials: es besteht aus mehreren Variablen (...), zwischen denen ein- bzw. wechselseitige Verbindungen existieren. Von „Problemlösen“ wird deshalb gesprochen, weil die Überführung des Ausgangszustands in einen (...) Zielzustand nicht auf Anhieb gelingt, sondern das Überwinden von „Barrieren“ zwischen Ist- und Sollwerten erfordert.“ (S.5).

Komplexe Problemlösesituationen sind nach Dörner et al. (1983) durch folgende fünf Merkmale charakterisiert:

- Komplexität
- Intransparenz
- Abhängigkeiten zwischen den beteiligten Variablen
- Eigendynamik
- Polytelie (Vielzieligkeit).

---

<sup>1</sup> Später untersuchten auch differentielle Psychologen detailliert die Steuerungsfertigkeiten von Individuen. Hierbei war das Ziel die Brauchbarkeit dynamischer Systemsteuerungsleistung als Prädiktor für den Berufserfolg im Rahmen der Eignungsdiagnostik zu untersuchen (z.B. Kersting, 1999). Parallel zu den psychologischen Forschungsgruppen beschäftigten sich auch Wirtschaftswissenschaftler wie die Gruppe um Serman (vgl. Serman, 1989) mit der Steuerungsleistung in dynamischen Systemszenarien.

Die Komplexität der dynamischen Steuerungssituation bezieht sich nach Dörner et al. (1983) auf die Menge der zu verarbeitenden Informationen. Diese überschreitet meistens die zur Verfügung stehende Kapazität und verhindert dadurch eine optimale Problemlösung. Unter Intransparenz versteht man, dass nicht alle entscheidungsrelevanten Informationen direkt zugänglich sind und somit eine aktive Informationsbeschaffung notwendig wird. Die (wechselseitigen) Abhängigkeiten zwischen den Systemvariablen müssen darüber hinaus erkannt werden. Ein weiteres Merkmal dynamischer Systeme ist ihre Eigendynamik, die sich darauf bezieht, dass sich die Situation auch ohne Zutun des Akteurs verändert. Dies kann Zeitdruck verursachen. Unter Polytelie versteht man nach Dörner et al. (1983) schließlich das häufige Vorliegen mehrerer, sich eventuell widersprechender Ziele. Dies erfordert eine mehrdimensionale Informationsbewertung und den Aufbau einer differenzierten Zielstruktur mit Regeln der Konfliktresolution. Funke (2001b) bewertet die Aspekte Vernetztheit und Dynamik als immanente, zentrale Systemmerkmale komplexer Probleme, die sich auf die Systemstruktur bzw. auf den zeitlichen Entwicklungsprozess der Systemparameter beziehen. Mit jedem Eingriff der steuernden Person in ein dynamisches System verändert sich der Ausgangszustand, der die Grundlage für den nächsten Eingriff darstellt. Dieser dynamisch an die Steuerungseingriffe angepasste Systemverlauf grenzt das komplexe Problem der dynamischen Systemsteuerung klar von statischen Urteils- oder Entscheidungsaufgaben ab, in denen die Elemente der Problemstellung unabhängig von den problemlösenden Personen sind.

Nach Steiner (1972, S.15) ist eine Aufgabe allgemein definiert als „a set of specifications identifying the goal that is to be achieved and the procedures that an individual or group may employ when attempting to achieve it“. Damit aus einem dynamischen System mit der Spezifikation von Vernetztheit und Dynamik und der entsprechenden Umsetzung in eine Computersimulation folglich eine *Systemsteuerungsaufgabe* resultiert, müssen das Steuerungsziel (bzw. detaillierte Teilziele), sowie die möglichen Steuerungseingriffe zur Zielerreichung definiert werden. Probanden in dynamischen Systemsteuerungsuntersuchungen wurden bisher mit Optimierungszielen wie beispielsweise der Wiederherstellung einer Optimaltemperatur in einem Kühlhaus (Reichert & Dörner, 1988), mit Maximierungszielen wie der Gewinnmaximierung in einem Händlerszenario (Kleiter, 1970) oder mit Minimierungszielen, wie der Minimierung von Toten und Infizierten in einem AIDS-Ausbreitungsszenario (Badke-Schaub, 1993) konfrontiert. Ungeachtet der jeweiligen semantischen Einkleidung stellen dynamische Systemsteuerungsaufgaben jedoch immer *Optimierungsaufgaben* dar, bei

denen es darum geht optimale Anpassungen der Eingriffsparameter vorzunehmen, um den aktuellen Systemzustand dem Zielzustand anzunähern. Die Möglichkeiten der Einflussnahme sind dabei durch eine definierte Anzahl von beeinflussbaren Systemparametern und durch programmierte Eingabeschnittstellen zur Systemsteuerung begrenzt. Die Rückmeldung über resultierende Systemzustände erfolgt in der Regel über die Ausgabeoberfläche der Computersimulation.<sup>2</sup>

Für die Fragestellung dieser Arbeit, die die Leistungsfähigkeit und den Lernzuwachs von Gruppen bei der Bearbeitung dynamischer Systemsteuerungsaufgaben fokussiert, ist eine Klassifikation des Aufgabentyps als Gruppenaufgabe von besonderer Relevanz. Stasser und Dietz-Uhler (2001) nennen zwei Dimensionen der Aufgabenklassifikation von Gruppenaufgaben, die sich sinnvoll auf dynamische Systemsteuerungsaufgaben anwenden lassen.<sup>3</sup> Die erste Dimension basiert auf dem Antwortformat einer Aufgabe und unterscheidet Selektionsaufgaben (selection tasks) von Ratingaufgaben (rating tasks). Während bei Selektionsaufgaben eine von mehreren Optionen zur Lösung der Aufgabe gewählt werden muss (z.B. die Auswahl des qualifiziertesten Bewerbers zur Besetzung einer vakanten Position), erfordern Ratingaufgaben eine quantitative Festlegung auf einer kontinuierlichen Antwortdimension (z.B. in einem Unternehmen die Festlegung der Produktionszahlen für den kommenden Monat). Dynamische Systemsteuerungsaufgaben sind den Ratingaufgaben zuzuordnen, da in der Regel quantitative Urteile zur Beeinflussung einer oder mehrerer Systemvariablen getroffen werden, um einen angestrebten Systemzustand zu erreichen oder beizubehalten. Häufig ist der quantitative Wertebereich der zu beeinflussenden Variablen dabei mit einem natürlichen Nullpunkt versehen und nach oben offen.

Die zweite Dimension zur Aufgabenklassifikation, die Stasser und Dietz-Uhler (2001) anführen, geht ursprünglich auf Laughlin (1980) zurück. Demnach kann zwischen intellektiven Aufgaben (intellective tasks) mit einer demonstrierbaren Lösung und Urteilsaufgaben (judgmental tasks) mit niedriger Demonstrierbarkeit der richtigen Lösung

---

<sup>2</sup> Bei der Mehrzahl der computersimulierten Systemsteuerungsaufgaben findet eine direkte Interaktion der Probanden mit dem Rechner statt (z.B. Reichert & Dörner, 1988; Diel & Sterman, 1995). Bei einigen Untersuchungen mit hoher Systemkomplexität und der Steuerung durch Gruppenurteile wie dem AIDS-Ausbreitungsszenario (Badke-Schaub, 1993) oder der Schneiderwerkstatt (Dörner, 1992) erfolgt eine auf die konkreten Nachfragen der Probanden beschränkte Rückmeldung der Systemreaktionen durch den Versuchsleiter.

<sup>3</sup> Obwohl Stasser und Dietz-Uhler (2001) ihr Kategorisierungsschema explizit auf Gruppenaufgaben beziehen, ist die Differenzierung von Selektions- und Ratingaufgaben ebenso bedeutsam für die Einordnung von Systemsteuerungsaufgaben, die von Individuen bearbeitet werden.

unterschieden werden. Laughlin und Ellis (1986) nehmen an, dass die meisten kognitiven Aufgaben auf einem Kontinuum zwischen rein intellektiven Aufgaben und reinen Urteilsaufgaben einzuordnen sind. Inwiefern eine Aufgabe nach Laughlin und Ellis gegenüber anderen Personen innerhalb einer Gruppe demonstrierbar ist, hängt von vier Bedingungen ab. Erstens müssen die Gruppenmitglieder bei der kollektiven Aufgabenbearbeitung ein konzeptuelles System teilen. Zweitens müssen sie Zugriff auf ausreichend Informationen haben, um auf die richtige Lösung schließen zu können. Drittens müssen die Individuen, die die korrekte Lösung gefunden haben, hinreichend fähig und motiviert sein, um diese plausibel zu erklären. Viertens müssen schließlich die Gruppenmitglieder, die die Aufgabe bisher nicht lösen konnten, die demonstrierte erfolgreiche Aufgabenbearbeitung verstehen und als richtig akzeptieren.

Dynamische Systemsteuerungsaufgaben sind aufgrund der Intransparenz zwischen den Systemvariablen und der häufig zeitverzögerten Systemreaktion auf Steuerungseingriffe nur begrenzt demonstrierbar. Selbst wenn die Zielvorgabe darin besteht, einen Optimalwert zu erreichen und ein kontinuierlich gewähltes quantitatives Steuerungsurteil die eindeutig richtige Lösung darstellt, wird diese Lösung nicht zu einem Steuerungszeitpunkt eindeutig demonstrierbar sein. In den meisten Systemen wird es zusätzlich erforderlich sein, über einen längeren Zeitraum mit mehreren Urteilszeitpunkten die richtige Steuerungsstrategie der Konstanthaltung der Optimaleinstellung zu verfolgen, damit sich der Systemzustand offensichtlich dem Zielzustand annähert und auf dem Zielwert stabilisiert. Da viele dynamische Systeme in ihrem Verlauf zusätzlich oszillatorischen Schwankungen oder Zufallsschwankungen unterliegen (Barlas & Dalkiran, 2008; Reichert & Dörner, 1988), wird sich der angestrebte Zielzustand nicht exakt erreichen lassen. Dies mindert unabhängig von der Steuerungsqualität die Demonstrierbarkeit. Folglich sind dynamische Systemsteuerungsaufgaben auf dem Klassifikationskontinuum von Laughlin und Ellis (1986) näher am Pol der Urteilsaufgaben lokalisiert. Je nach Ausgestaltung des dynamischen Systems ist die Demonstrierbarkeit einer richtigen Steuerung bzw. einer Steuerungsstrategie dabei mehr oder weniger stark eingeschränkt. Hinzu kommt, dass sich die Systemparameter entsprechend vorausgehender Eingriffe im Steuerungsverlauf dynamisch verändern und somit je nach Verlauf und Zeitpunkt spezifische Steuerungsurteile zielführender und leichter als geeignet zu demonstrieren sein können als der bei optimalem Steuerungsverlauf favorisierte Steuerungsinput. Zusammenfassend lässt sich für dynamische Steuerungsaufgaben nach dem Kategorisierungsschema von Stasser und Dietz-Uhler (2001), das in Tabelle 1 dargestellt ist, eine Zuordnung zu der Kategorie *Judgment Tasks* vornehmen, deren Bearbeitung durch

ein quantitatives Rating und eine niedrige bis mittlere Demonstrierbarkeit der Urteilsaufgabe geprägt ist. Damit widerspricht diese Aufgabenkategorisierung der Begriffswahl des „komplexen Problems“ die Dörner et al. (1983) für dynamische Steuerungsaufgaben wählten. Entsprechend der Klassifikation von Stasser und Dietz-Uhler werden dynamische Systemsteuerungsaufgaben in dieser Arbeit konsistent als *Urteilsaufgaben* und Systemeingriffe entsprechend als *Steuerungsurteile* bezeichnet.<sup>4</sup> Damit erfolgt gleichzeitig eine klare Abgrenzung zu *statischen* Aufgaben der Problemlöseforschung, deren Aufgabenstruktur unabhängig vom Verhalten des Problemlösers ist und sich nicht dynamisch im Zeitverlauf verändert.<sup>5</sup>

Tabelle 1: *Schema zur Kategorisierung von Gruppenaufgaben nach Stasser und Dietz-Uhler (2001, S. 34)*

	<i>Response Format</i>	
	<i>Select</i>	<i>Rate</i>
<i>Judgmental</i>	Choice	Judgment
<i>Intellective</i>	Problemsolving	Estimation

## 2.2 Forschungsstand zur Steuerung dynamischer Systeme

Die hohe praktische Relevanz von dynamischen Systemsteuerungsaufgaben hat dazu geführt, dass seit den 1970er Jahren zahlreiche Studien zur menschlichen Steuerungsleistung und zur Aneignung von dynamischem Systemwissen aus dem Blickwinkel unterschiedlicher Forschungsdisziplinen wie beispielsweise den Wirtschaftswissenschaften (Serman, 1994), der Psychologie (Broadbent, FitzGerald, &

<sup>4</sup> Viele Publikationen, insbesondere von Autoren mit wirtschaftswissenschaftlichem Hintergrund, verwenden die Bezeichnung *Dynamic Decision Making*, also dynamische Entscheidungsaufgaben (z.B. Barlas & Dalkiran; 208; Brehmer, 1992; Diehl, 1992; Diehl & Serman, 1995; Hsiao, 2000; Serman, 1989a). Das Konzept *Decision Making* ist nicht eindeutig definiert und umfasst je nach Forschungsgegenstand die Konzepte *Choice* (Wahlaufgaben) und/oder auch *Judgment* (Urteilsaufgaben). Aufgrund dieser begrifflichen Unschärfe wird es in dieser Arbeit nicht verwendet.

<sup>5</sup> Beispiele für statische Problemlöseaufgaben sind die Wasserumfüllaufgabe von Atwood und Polson (1976) und das Zwei-Seile-Problem von Maier (1931).



Broadbent, 1986) und der Ergonomie (Howie, Sy, Ford, & Vicente, 2000) resultierten. Die Studien verwenden Computersimulationen von dynamischen Systemen, deren Komplexität von zwei Systemvariablen in einfachen Regelkreisen (Reichert & Dörner, 1988) bis zu über 2000 Variablen bei der Simulation einer Kleinstadt (Dörner et al., 1983) variiert. Eine Reihe von Reviews stellt die zahlreichen bisherigen Befunde zur Steuerungsleistung in der Übersicht dar (vgl. Gonzalez, Vanyukov, & Martin, 2005; Hsiao & Richardson, 1999; Lipshitz, Klein, Orasanu, & Salas, 2001; Maier & Größler, 2000; Meadows, 2007; Osman, 2010; Rouwette, Größler, & Vennix, 2004). Im folgenden Abschnitt wird beschrieben, welche Beurteilungsverfahren zur Steuerungsleistung von Individuen in der Forschung eingesetzt werden und zu welchen zentralen Ergebnissen die Simulationsstudien bei der Untersuchung der menschlichen Systemsteuerungsleistung und des Erwerbs von Systemwissen kommen.

### **2.2.1 Systemsteuerungsleistung und systembezogener Wissenserwerb von Individuen**

In bisherigen Arbeiten zur Leistung bei der Steuerung dynamischer Systeme werden vielfältige Beurteilungskriterien der Steuerungsqualität mit unterschiedlicher Objektivität herangezogen. Für Steuerungsszenarien, in denen die quantitative Ausprägung einer Outputvariablen leistungsentscheidend ist<sup>6</sup>, haben sich insbesondere die drei Beurteilungsverfahren *Expertenrating*, *Sollwertabweichung* und *Vergleich mit Benchmarks* etabliert.

Das Expertenrating stellt ein Beurteilungsverfahren mit relativ geringer Objektivität dar. Hierbei beurteilen Rater, die in Bezug auf das dynamische System und mögliche Systemverläufe geschult sind, nach Abschluss der Steuerungsphase die Steuerungsgüte. Grundlage der Beurteilung können sowohl Endergebnisse der Steuerungsphase als auch der Gesamtverlauf der Steuerung sein. Meistens erfolgt eine kategoriale

---

<sup>6</sup> In sehr komplexen Szenarien mit mehreren leistungsrelevanten Outputvariablen erfolgt häufig zuerst die separate Einstufung der Outputvariablen auf einer Leistungsskala und anschließend eine Integration der Teilleistungen zu einem Gesamtleistungswert (z.B. Broadbent & Ashton, 1978). Bei anderen Vorgehensweisen müssen die Systemwerte mehrerer Bewertungskriterien jeweils innerhalb eines bestimmten Wertebereichs liegen, damit eine Gesamtbewertungskategorie zutrifft. So muss z.B. in der AIDS-Ausbreitungssimulation von Badke-Schaub (1993) die höchste Anzahl Infizierter, die höchste Anzahl aktuell Erkrankter und die kumulierte Anzahl von AIDS-Fällen jeweils unterhalb eines bestimmten Grenzwertes liegen, damit die Gesamtsteuerungsleistung mit 1 (sehr gut) bewertet wird.

Leistungseinstufung wie beispielsweise bei Reichert und Dörner (1988) auf einer fünfstufigen Skala von 1 (schlechte Steuerung) bis 5 (erfolgreiche Steuerung).

Das zweite Verfahren der Berechnung der Sollwertabweichung ermöglicht eine objektive, quantitative Bestimmung der Steuerungsleistung. Es bietet sich an, wenn das Steuerungsziel darin besteht, einen Optimalzustand im System herbeizuführen. Hierbei werden die Abweichungen der durch die Steuerung erzielten Systemzustände vom Sollwert für jeden Zeittakt der Steuerung berechnet und anschließend aufsummiert, um die Gesamtleistung in einem bestimmten Steuerungszeitraum abzubilden (z.B. Reichert & Dörner, 1988).

Als drittes Verfahren zur objektiven Beurteilung des absoluten Leistungsniveaus werden oftmals so genannte Benchmarks verwendet. Benchmarks sind Vergleichsmaßstäbe im möglichen Leistungsspektrum der Systemsteuerung innerhalb des simulierten Systems. Als Vergleichsmaßstab für eine schlechte Steuerungsleistung wird die *no control rule* eingesetzt (z.B. Diehl, 1992). Hierbei wird die Ausprägung der leistungsrelevanten Systemvariablen bestimmt, die aus der eigendynamischen Entwicklung eines Systems ohne steuernde Eingriffe resultiert. Diese Systementwicklung ohne Eingriffe kann als plausible Untergrenze für die zielbezogene Steuerungsleistung angesehen werden.<sup>7</sup> Als Vergleichsmaßstab für eine sehr gute Steuerungsleistung werden entweder Steuerungsergebnisse herangezogen, die aus der Anwendung einer optimalen Eingriffsstrategie über den gesamten Steuerungszeitraum resultieren (z.B. Reichert & Dörner, 1988) oder es wird post hoc mathematisch zu jedem Steuerungsdurchgang aus dem aktuellen Systemzustand der optimale Eingriffswert bestimmt (z.B. Diehl & Sterman, 1995)<sup>8</sup>. Die drei beschriebenen Verfahren Expertenrating, Sollwert-Abweichungsberechnung und Vergleich mit Benchmarks können sowohl zur Bewertung der absoluten Systemsteuerungsleistung von Individuen als auch zur absoluten Leistungsbewertung von Gruppen eingesetzt werden.

Zahlreiche Forschungsbefunde zur individuellen Systemsteuerung unter den genannten Leistungsbeurteilungskriterien belegen, dass Probanden Defizite im Umgang

---

<sup>7</sup> Die aus der *no control rule* resultierenden Systemzustände stellen in der Regel nicht das absolut schlechteste Ergebnis in Bezug auf das Steuerungsziel dar. Durch kontraproduktive Systemeingriffe kann bei den meisten Systemen eine noch größere Diskrepanz zum Zielkriterium hervorgerufen werden, die dann der absoluten Untergrenze des Steuerungsspektrums entspricht.

<sup>8</sup> Bei dieser dynamisch an den Systemverlauf angepassten Bestimmung der optimalen Steuerungsurteile kommen in den meisten Fällen Methoden des Operations- Research zum Einsatz (vgl. Petzing, 1993).

mit dynamischen Systemen aufweisen und bei ihrem Handeln die Eigendynamik des Systems, die zeitverzögerten Folgen und die Nebeneffekte ihrer Handlungen sowie die Nonlinearität von Zusammenhängen nicht ausreichend berücksichtigen (z.B. Lee, Padmanabhan, & Whang, 1997; Paich & Serman, 1993; Serman, 1989b; Serman, 1994). Diese Fehlwahrnehmung des Systemfeedbacks auf die eigenen Steuerungseingriffe wird als *misperceptions of feedback hypothesis* bezeichnet (vgl. Diehl & Serman, 1995; Serman, 1989a) und führte zu deutlich höheren Kosten als bei optimaler Steuerung zu erwarten gewesen wären. Beispielsweise verursachten die Probanden im beer distribution game, einer dynamischen Warenhandelskette, 10mal höhere Kosten als bei Nutzung der optimalen Steuerungsstrategie (Serman, 1989b). Die beobachtbaren systematischen Schwächen in der individuellen Steuerungsleistung zeigten sich sowohl bei abstrakten Systemen ohne realitätsnahe semantische Einkleidung der Variablenbezeichnungen (Funke, 1992) als auch bei der Steuerung realitätsnaher Wirtschaftsszenarien durch betriebswirtschaftliche Laien und Experten (Lee, Padmanabhan, & Whang, 1997).

Nach Funke (2001a) stellt die erfolgreiche Steuerung von dynamischen Systemen zwei Aufgabenanforderungen an die steuernden Personen: Wissenserwerb bezüglich der Beschaffenheit des Systems und Wissensanwendung im Sinne einer zielführenden Systemsteuerung. Kluge (2004) schlussfolgert nach einer systematischen Aufarbeitung der empirischen Forschungsergebnisse in komplexen Systemsteuerungsszenarien: „Hinsichtlich des Zusammenhangs von Wissen und Steuerleistung kommen die meisten Studien zu einem positiven Ergebnis“ (S. 116). Die Befundlage zur Steigerung des Systemwissens durch freie Systemexploration bzw. Steuerungserfahrung ist gemischt. Während einige Studien eine Verbesserung des individuellen Verständnisses von Variablenzusammenhängen und weiteren Systemeigenschaften zeigten (z.B. Kröner, 2001; Putz-Osterloh, Bott, & Houben, 1988), fanden andere Studien keine Wissenssteigerung durch Systemerfahrung (z.B. Funke & Müller, 1988; Preußler, 1997). Bei sehr komplexen Systemen wirkt sich eine gestufte Systemkonfrontation positiv auf die Leistung und das explizite Systemwissen aus (vgl. Broadbent, FitzGerald, & Broadbent, 1986). Hierbei erfolgt eine angeleitete Exploration mit Steuerungseingaben von Teilsystemen, bevor das Gesamtsystem gesteuert werden muss. Nach Kluge (2004) hat eine angeleitete Systemexploration (z.B. mit dem Hinweis, eine Variable zu einem Zeitpunkt zu manipulieren und systematisch Systemreaktionen zu beobachten) häufiger einen positiveren Einfluss auf den Wissenserwerb als die freie Systemexploration ohne Vorgaben.

Ob eine gegensätzliche Kausalbeziehung vorliegt und umfangreicheres bzw. korrekteres Systemwissen zu einer gesteigerten Steuerungsleistung führt, wurde mehrfach durch experimentell induziertes Strukturwissen überprüft. Hierbei wurde der Experimentalgruppe Strukturwissen über das System, beispielsweise in Form von grafischen Kausaldiagrammen (z.B. Kieras & Bovair, 1984; Süß, 1996), in Informationstexten oder einer Kombination aus beidem dargeboten. Anschließend wurde die Steuerungsleistung mit einer Kontrollgruppe ohne Systemstrukturinformation verglichen. Dabei zeigt die Mehrzahl der Studien keinen signifikanten leistungssteigernden Effekt durch vorgegebenes Strukturwissen auf die Steuerungsleistung (Berry & Broadbent, 1984; Heineken, Arnold, Kopp, & Soltysiak, 1992; Preußler, 1996; Putz-Osterloh, 1993; Süß, 1996). Ausnahmen stellen die Studien von Kröner (2001) sowie von Kieras und Bovair (1984) dar.

Vollständige Informationen über das Simulationsmodell in der Studie von Kröner (2001) in Form eines Kausaldiagramms führten allerdings nicht generell, sondern nur bei den intelligenteren Probanden der Stichprobe zu einem Anstieg der Steuerungsleistung. In der Studie von Kieras und Bovair (1984) wurde das zur Verfügung gestellte Systemwissen durch einen Wissenstest mit sofortigem Feedback für die Probanden ergänzt, dessen Fragen sie wiederholt beantworten mussten, bis alle Antworten richtig waren. Demnach wurde in dieser Studie von den Autoren vollständig korrekt repräsentiertes Systemwissen bei den Probanden sichergestellt. Das erklärt den positiven Effekt auf die Steuerungsleistung, der in den oben genannten weiteren Studien nicht auftrat. Dort wurde Wissen zur Verfügung gestellt, aber die richtige Wissensrepräsentation vor der Steuerung nicht überprüft bzw. veranlasst. Berry und Broadbent (1984) zeigten einen signifikanten Interaktionseffekt von Steuerungsleistung und Instruktion zum lauten Denken: Probanden mit experimentell induziertem Systemwissen zeigten dann Leistungssteigerungen, wenn sie während der Steuerung explizit ihre Steuerungsurteile begründen mussten und somit bewusst Bezug auf ihr Systemwissen nahmen. Kluge (2004) wies experimentell nach, dass eine Interaktion von Lernform und Schwierigkeitsgrad der Systemsteuerung auftreten kann und die Vermittlung von Strukturwissen gegenüber der Kontrollbedingung mit Systemexploration keinen positiven Leistungseffekt unter geringer Schwierigkeit, sondern erst bei „mittlerer“ Systemschwierigkeit aufweist. Demnach stellt die Systemkomplexität eine mögliche Moderatorvariable dar, die möglicherweise die heterogenen Befunde unterschiedlicher Studien zum Zusammenhang von Wissen und Steuerungsleistung erklären kann.

Angesichts der negativen gesamtgesellschaftlichen Konsequenzen, die individuelle Steuerungsdefizite von Entscheidungsträgern im politischen, wirtschaftlichen und sicherheitstechnischen Bereich nach sich ziehen können, scheint es dringend geboten, leistungsförderliche Faktoren zu identifizieren und in der Praxis zu implementieren. Die oben erläuterten, empirisch auf ihre Wirksamkeit untersuchten Interventionen zur Verbesserung des *individuellen* Systemverständnisses und der *individuellen* Steuerungsleistung durch Zugang zu graphisch oder numerisch aufbereitetem Wissen über Systemstrukturen sowie durch die angeleitete Exploration stellen dabei keine zufriedenstellende Lösung dar: Maßnahmen zur Steigerung des theoretischen Systemwissens und der Steuerungskompetenz von steuerungsunerfahrenen Personen (vgl. für einen Überblick Kluge, 2004, S. 94) sowie von systembezogenen Experten (z.B. Serman, 1989b) erzielten zwar positive Effekte, ließen aber deutlichen Spielraum für Steigerungen.

An diesem Punkt setzt die Fragestellung dieser Arbeit an, die sich auf die möglichen leistungsförderlichen Effekte der Systemsteuerung von Gruppen (im Vergleich zur individuellen Steuerungsleistung) bezieht und aus zwei Komponenten besteht:

1. Kann eine Steuerung durch Gruppen die vorhandenen individuellen Defizite in der Steuerungsleistung und im Systemverständnis kompensieren?
2. Kann die Gruppen*interaktion* darüber hinaus zu einer besseren Steuerungsleistung führen als durch den Einsatz vergleichbarer personaler Ressourcen bei individueller Steuerung erzielt werden kann?

Im folgenden Abschnitt werden zunächst bisherige Forschungsergebnisse zur Systemsteuerungsleistung und zum Wissenserwerb von Gruppen berichtet und mit methodischen Mängeln einhergehende Forschungslücken aufgezeigt. Anschließend erfolgt die Beschreibung relevanter sozialpsychologischer Theorien und Forschungsmethoden zur Untersuchung der Fragestellung und die Darstellung des Nutzens, der sich aus der Verknüpfung der sozialpsychologischen Gruppenforschung mit dem Paradigma der dynamischen Systemsteuerung für den theoretischen und praktischen Erkenntnisgewinn ergibt.

### **2.2.2 Systemsteuerungsleistung und systembezogener Wissenserwerb von Gruppen**

Zur dynamischen Systemsteuerungsleistung im Gruppenkontext liegen weitaus weniger Studien vor als zur Steuerungsleistung von Individuen. Für die Beurteilung der

Gruppenleistung können einerseits die gleichen absoluten Benchmarks wie für die individuelle Leistungsbeurteilung verwendet werden. Andererseits ist es für den Personaleinsatz im beruflichen Kontext interessant zu überprüfen, ob Gruppen in bestimmten Aufgabenszenarien in Relation zu Individuen eine bessere Steuerungsleistung erbringen. Die bisher in der Literatur berichteten Studien erlauben jedoch aus unterschiedlichen Gründen keinen zuverlässigen Rückschluss auf die Unter- bzw. Überlegenheit von Gruppen.

Die Studien von Dörner (1992) sowie von Broadbent und Ashton (1978) untersuchten ausschließlich Gruppen bei der dynamischen Systemsteuerung, ohne einen direkten Vergleich zur Steuerungsleistung von Individuen vorzunehmen. Dörner (1992) beschrieb in narrativer Form den Interaktionsprozess und die Steuerungsurteile von Dreier- und Vierergruppen bei der Steuerung einer simulierten Schokoladenfabrik. Er beobachtete in Einzelfallanalysen die Missachtung von Neben- und Fernwirkungen vorgenommener Eingriffe (wie sie auch bei der individuellen Steuerung auftreten) und wies auf Urteilsfehler der Gruppen hin. Broadbent und Ashton (1978) analysierten die Leistung von 14 Gruppen mit durchschnittlich 5 Mitgliedern, die in einer Simulation des britischen Wirtschaftssystems die Zielvorgabe hatten, den Wohlstand zu maximieren. Die Studie zeigte auf, dass eine deutliche Leistungsvariabilität der Gruppen innerhalb der Stichprobe auftrat.<sup>9</sup>

Wenn man positive Effekte des Einsatzes von Gruppen bei der Steuerung dynamischer Systeme im Vergleich zum Einsatz individueller Steuerung betrachten möchte, können die folgenden zwei Fragestellungen unterschieden werden:

1. Ist eine Gruppe bei der Systemsteuerung besser als ein Individuum? Sind also  $N$  Köpfe besser als einer?
2. Profitieren  $N$  Personen davon, wenn sie bei der Systemsteuerung *gemeinsam* Steuerungsurteile fällen, oder könnte man sich die Gruppeninteraktion im Hinblick auf die Steuerungsleistung auch sparen? Sind also  $N$  Köpfe, die in der Gruppe zusammenarbeiten, besser als  $N$  getrennt urteilende Köpfe?

---

<sup>9</sup> Außerdem untersuchten Broadbent und Ashton (1978) mit einem Wissensfragebogen zu Variablenzusammenhängen, ob die Steuerungserfahrung einen Effekt auf das Systemwissen der Gruppenmitglieder hatte. Insgesamt konnten sie keine Zunahme des Systemwissens feststellen. Dieser Befund ist allerdings nur unter Vorbehalt zu nennen. Die Rücklaufquote des Wissensfragebogens umfasste weniger als 50% der Stichprobe und es ist denkbar, dass sich die Non-Responder systematisch in ihrem Systemwissen von den Respondern unterschieden und die Auswertung der Gesamtstichprobe zu einem anderen Ergebnis geführt hätte.

Die bisher in der Literatur berichteten Studien fokussieren ausschließlich die erste Fragestellung und vergleichen die Leistung von Gruppen mit der Steuerungsleistung einer (annähernd) gleichen Anzahl von Individuen. Bei diesen Studien liegt generell eine größere kognitive Gesamtkapazität in der Gruppenbedingung als in der individuellen Vergleichsbedingung vor. Die größeren kognitiven Ressourcen könnten eine Überlegenheit der Gruppen gegenüber den Individuen begünstigen. Mit einer Ausnahme (Wolfe & Chacko, 1983) kamen jedoch alle betrachteten Studien zu dem Ergebnis, dass sich die Gruppenleistung nicht signifikant von der individuellen Systemsteuerungsleistung unterschied (Badke-Schaub, 1993; Dentler, 1977; Endres & Putz-Osterloh, 1994; Köller, Dauheimer, & Strauß, 1993; Leutner, 1988; Putz-Osterloh, 1991). Die genannten Studien weisen größtenteils spezifische methodische Mängel auf. Diese beziehen sich, wie nachfolgend detailliert erläutert wird, auf experimentelle Konfundierungen und somit auf einen unfairen Vergleich zwischen der Gruppen- und der Individualbedingung. Die Studien liefern daher keine eindeutigen Erkenntnisse.

Wolfe und Chacko (1983) ließen Studierende der Wirtschaftswissenschaften über den Verlauf eines Semesters in 8 Runden eine komplexe Wirtschaftssimulation steuern und verglichen die individuelle Steuerungsleistung mit der Leistung von Gruppen mit 2 bis 4 Mitspielern. Gruppen jeder untersuchten Größe erzielten dabei eine bessere Leistung mit signifikant höherem Gewinn als individuell steuernde Probanden. Dreiergruppen erzielten gegenüber Dyaden eine signifikante Leistungssteigerung. Vierergruppen entsprachen in der Leistung den Gruppen mit drei Mitgliedern. Diese Studie ist die einzige, die darauf hindeutet, dass das auf Gruppenurteilen basierende Steuerungsoutput dem individuellen Steuerungsoutput überlegen ist. Besonderheiten der Aufgabenstruktur und des Kontextes der Datenerhebung erlauben allerdings nicht die eindeutige Interpretation, dass es „bessere“ Gruppensteuerungsurteile per se waren, die zur Überlegenheit der Gruppen führten. Die Aufgabe ermöglichte 56 Urteilsoptionen pro Urteilsdurchgang und beinhaltete dabei die Besonderheit, dass die Probanden aktiv Informationen einholen bzw. über den Kauf von Informationen entscheiden konnten. Individuen holten durchschnittlich weniger Informationen ein als die Steuerungsgruppen. Als erster methodischer Mangel dieser Studie ist somit die systematisch kleinere Informationsbasis der Individuen für das Treffen der Steuerungsurteile im Vergleich zu den Gruppen zu sehen.<sup>10</sup> Es ist denkbar, dass die Individuen bei Ausgleich ihres Informationsnachteils durch Konstanthaltung der

---

<sup>10</sup> Da die Bearbeitungszeit begrenzt war, waren die Gruppen möglicherweise zusätzlich aufgrund ihrer größeren kognitiven Kapazität im Vorteil und konnten die Informationssichtung als teilbare Subaufgabe der Systemsteuerung effizienter erledigen.

verfügbaren Informationen über alle Bedingungen vergleichbar gute Steuerungsergebnisse erzielt hätten wie die Gruppen. Ein weiteres methodisches Problem stellt die unkontrollierte Wissensaneignung durch die Bearbeitung von zusätzlichen Fallstudien und den Besuch von Vorlesungen im Rahmen des Kurses über den Verlauf des Semesters da. Es ist unklar, inwiefern die Studierenden in allen Bedingungen vollständige bzw. vergleichbare Anwesenheit bei den ergänzenden Lehrveranstaltungen zeigten. Während Individuen eventuelle Fehlzeiten schlecht kompensieren konnten, bestand im Gruppenkontext die Möglichkeit, verpasste Inhalte auszutauschen und somit das gesamte vermittelte Wissen des Wirtschaftskurses als Steuerungsgrundlage zur Verfügung zu haben. Darüber hinaus wurde nicht untersucht, ob die kooperative Aufarbeitung der simulationsexternen Lehrinhalte in den Steuerungsgruppen einen leistungssteigernden Effekt hatte, von dem die Individuen nicht profitieren konnten. Studien zum kollaborativen Lernen zeigten entsprechend, dass Mitglieder von kooperativen Lerngruppen mehr Wissen erwerben und bessere Transferleistungen erbringen als Lernende, die sich alleine mit dem präsentierten Unterrichtsstoff auseinandersetzten (vgl. für einen Überblick Johnson, Johnson, & Stanne, 2000). Somit könnte in der Studie von Wolfe und Chacko (1983) eine zweite Konfundierung zwischen Gruppenkontext und steuerungsrelevanter Informationsbasis vorliegen, die sich auf die vollständige Verfügbarkeit und Repräsentation simulationsexterner Lehrinhalte bezieht.

Die Studien von Dentler (1977) zu einem Ökosystem und von Badke-Schaub (1993) zu einem AIDS-Ausbreitungsszenario verglichen jeweils Dreiergruppen mit einer annähernd vergleichbaren Anzahl an Probanden, die individuell das System steuerten. Dentler kam zu dem Ergebnis, dass sich die Gruppenleistung nicht von der Individualleistung unterschied. Diese Ergebnisinterpretation kann aus drei Gründen nicht als gesichert angesehen werden. Erstens war das Steuerungsziel, das sich auf die Verbesserung der Lebensbedingungen in einem fiktiven afrikanischen Staat bezog, offen formuliert und musste von den Probanden subjektiv konkretisiert werden. Entsprechend gab es keine objektiven Leistungskriterien. Dentler (1977) definierte post hoc die Menge an Eingriffen und die Verlaufsstabilität einer Reihe von Systemvariablen als Leistungsindikatoren. Hieran ist problematisch, dass die subjektiven Leistungskriterien der Probanden (beispielsweise festgelegte Maximierungsziele bzgl. der Variablenausprägung) davon abweichen konnten und der Steuerungserfolg nicht in Relation zur Zielinterpretation gemessen wurde. Zweitens wurden die von Dentler (1977) erhobenen Gruppendaten mit einem Steuerungshorizont von 200 Monaten mit den individuellen Steuerungsdaten aus einer anderen Studie (Dörner & Reither, 1978) mit einem Planungshorizont von nur 130



Monaten verglichen. Drittens arbeiteten die Probanden in der Individualbedingung streng genommen nicht allein, sondern es bestand eine dyadische Interaktionssituation mit dem Versuchsleiter, in der fortwährend Informationen über das System ausgetauscht wurden und nicht auszuschließen ist, dass konstruktives Probandenverhalten wie Reflexion bzw. „lautes Denken“ durch den Versuchsleiter bekräftigt wurde (vgl. Dentler, 1977, S.198).

In der Studie von Badke-Schaub (1993) hatten die Einzelpersonen bei der simulierten AIDS-Ausbreitungsbekämpfung in einer Großstadt konstant die Rolle eines bestimmten Interessenvertreters (des Bürgermeisters) zugewiesen bekommen, wohingegen innerhalb der Gruppen drei unterschiedliche Rollen zugewiesen wurden und Eingriffe aus der Perspektive des Bürgermeisters, des Gesundheitsamtdirektors und des Minderheitenvertreters für Homosexuelle geplant wurden. Somit gab es eine experimentell induzierte Konfundierung der Steuerungsbedingung mit der Sichtweise der Szenariobetrachtung, die einen Vorteil für die Dreiergruppen darstellte. Trotz der Vorteile der größeren kognitiven Ressourcen (und der Multiperspektivität im AIDS-Szenario) erzielten die Gruppen in dieser Studie ebenfalls keinen Steuerungsvorteil gegenüber den individuell steuernden Probanden.

In drei weiteren Studien versuchten die jeweiligen Autoren ein vergleichbares kognitives Leistungslevel zwischen der Gruppen- und der Individualbedingung im Vorfeld der Systemsteuerung sicherzustellen. Hierzu wurden die Probanden gemäß ihres Intelligenztestergebnisses (Berkowitz & Szabo, 1978) oder gemäß der individuellen Steuerungsleistung in einem anderen, nicht in der Hauptsteuerungsphase verwendeten Systemszenario (Köller, Dauheimer, & Strauß, 1993; Putz-Osterloh, 1991) in homogene Dyaden mit entweder hoher oder niedriger Ausgangsleistung beider Dyadenmitglieder eingeteilt und mit Einzelpersonen mit entsprechend hoher oder niedriger Ausgangsleistung verglichen. Dieses Versuchsdesign eliminierte dadurch nicht „das Problem“, dass die kognitiven Ressourcen in der Dyadenbedingung größer waren als in der Individualbedingung und brachte für jede der drei Studien ein weiteres spezifisches methodisches Problem mit sich. Der von Berkowitz und Szabo (1978) verwendete Intelligenztest ist nach neueren Erkenntnissen kein geeigneter Prädiktor für die Problemlösefähigkeit (Gebauer & Mackintosh, 2007; Kanfer & Ackerman, 1989; Rigas & Brehmer, 1999). Daher war nicht sichergestellt, dass das Ausgangslevel der Dyaden dem der für die Auswertung zugeordneten Individuen entsprach. Putz-Osterloh ließ Dyaden in ihrer Studie von 1991 (zitiert nach Köller et al., 1993) zwei Szenarien steuern, die individuell steuernden Probanden wurden hingegen nur mit einem Szenario konfrontiert. Somit lag eine Konfundierung mit dem Ausmaß an Steuerungserfahrung vor, das die individuellen

Steuerer benachteiligte. Das Experiment von Köller und Kollegen (1993) nutzte die Steuerungsleistung in einem Ökosystem, um Probanden als hoch oder niedrig kompetent zu klassifizieren und anschließend allein oder im Dyadenkontext ein simuliertes Wirtschaftssystem steuern zu lassen. Andere Befunde zeigen, dass die Steuerungsleistung interindividuell mit dem semantischen Kontext eines Systems variieren kann (z.B. Bakken, 1993), sodass auch hier die Validität der Messung der Ausgangskompetenz fraglich ist.

Zusammenfassend verwendeten also bisherige Studien den Vergleich von  $N$  Gruppen mit  $N$  Individuen, um Aussagen über die relative Steuerungsleistung von Gruppen bei der Steuerung dynamischer Systeme zu treffen. Dieses Design bevorzugt die Gruppenbedingung, in der  $n$ -mal so große kognitive Kapazitäten zur Verfügung stehen und die Gruppen auch von einem statistischen Aggregationsvorteil profitieren können, beispielsweise durch Fehlerausgleich bei Mittelung mehrerer Fehler behafteter Individualurteile.<sup>11</sup> Trotzdem unterschied sich die Gruppensteuerungsleistung in der Mehrzahl der Studien nicht von der individuellen Leistung. Die Studie von Wolfe und Chacko (1983) liefert einen Hinweis darauf, dass die Gruppensteuerung in einem hochkomplexen Wirtschaftsszenario einen signifikanten Leistungsvorteil erzielen kann. Aus methodischen Gründen kann die festgestellte Überlegenheit der Gruppen aber nicht eindeutig auf qualitativ höherwertige Entscheidungen bzw. Steuerungsurteile zurückgeführt werden, da sie mit der zugrundeliegenden Systeminformationsmenge konfundiert war und nicht für kollaboratives Lernen ergänzender Lehrinhalte kontrolliert wurde. Somit ist die erste genannte Fragestellung, ob Gruppen bessere Steuerungsleistungen bei der Kontrolle dynamischer Systeme erbringen können als *Einzelpersonen* bisher noch nicht unter Kontrolle bzw. Konstanthaltung leistungsrelevanter Kontextfaktoren untersucht worden.

Die zweite Fragestellung soll in dieser Arbeit erstmals fokussiert werden. Hierbei wird untersucht, ob Gruppen aufgrund der häufig zeitintensiven *Interaktion* Leistungssteigerungen bei der Systemsteuerung gegenüber derselben Anzahl von Personen bei *vergleichbarer individueller Steuerung* erzielen. In Kapitel 2.3.1 wird erläutert, mit welchem methodischen Vorgehen untersucht werden kann, ob Gruppeninteraktion per se leistungsförderliche Effekte bei der dynamischen Systemsteuerung hervorruft, die über die Effekte der kognitiven Gesamtkapazität und des statistischen Aggregationsvorteils der Individualurteile in der Gruppe hinausgehen.

---

<sup>11</sup> In Abschnitt 2.3.1 wird genauer auf die Problematik eingegangen, inwiefern zwischen einem statistisch bedingten Leistungsvorteil der Gruppen und einem genuin interaktionsspezifischen Leistungsvorteil von Gruppen gegenüber Individuen differenziert werden kann.

Der Einfluss von Gruppeninteraktion auf das Systemwissen wurde in einer Studie von Meyer und Scholl (2009) untersucht. Im Wirtschaftsszenario einer Schneiderwerkstatt erfassten die Autoren den Einfluss einer fünfzehnminütigen unstrukturierten Diskussion in Dyaden auf das Systemwissen und die Steuerungsleistung der Probanden. Die Autoren manipulierten das vor der Diskussion vorhandene Systemwissen und konnten zeigen, dass ein umgekehrt u-förmiger Zusammenhang zwischen dem Überschneidungsausmaß der Wissensrepräsentationen der Probanden und dem Ausmaß an Wissenszuwachs (lediglich deskriptiv mit  $p = .086$ ) und Steuerungsleistung bestand. Das Ergebnis kann so interpretiert werden, dass sich die Gruppeninteraktion dann besonders positiv auf den Wissenserwerb auswirkt, wenn eine gemeinsame Wissensbasis (im Sinne gegebener Demonstrierbarkeit ausgetauschter Informationen) vorhanden war. Wiesen die Wissensrepräsentationen der Dyadenmitglieder keine Gemeinsamkeiten auf oder waren identisch, resultierte hingegen kein individueller Wissenszuwachs aus der Gruppendiskussion.<sup>12</sup> Wolfe und Chacko (1983) verglichen den individuellen Lernzuwachs von Studierenden, die entweder alleine, in Dyaden, in Dreier- oder in Vierergruppen über mehrere Wochen ein komplexes Wirtschaftssimulationssystem steuerten. Der größte individuelle Wissenszuwachs resultierte dabei in Steuerungsgruppen mit 3 oder 4 Mitgliedern. Dies Ergebnis kann entsprechend des Befundes von Meyer und Scholl (2009) so interpretiert werden, dass es mit steigender Gruppengröße wahrscheinlicher ist, innerhalb der Gruppe heterogenes Systemwissen vorzufinden, das wechselseitiges Dazulernen ermöglicht. Wie bereits oben zur Studie von Wolfe und Chacko (1983) erläutert wurde, holten die Gruppen aktiv mehr neues Wissen während der Steuerung ein. Daher ist unklar, ob der individuelle Wissenszuwachs auch unter Konstanzhaltung der Menge des verfügbaren Systemwissens für Individuen und Gruppen zu einem größeren Lernzuwachs im Gruppenkontext geführt hätte. Keine der beiden beschriebenen Studien zum Wissenserwerb kann also die Frage beantworten, ob die interaktive Steuerung im Gruppenkontext im Vergleich zur individuellen Steuerung (und unter Konstanzhaltung der zur Verfügung gestellten expliziten Systeminformationen) zu einer signifikanten Steigerung des Systemwissens- bzw. Systemverständnisses führt. Auch an dieser Forschungslücke soll in dieser Arbeit angesetzt werden.

---

<sup>12</sup> Das Experiment von Meyer und Scholl (2009) beinhaltete keine Kontrollbedingung ohne Dyadeninteraktion, sondern es wurde lediglich der Einfluss der unabhängigen Variablen „Informationsüberlappung“ in den drei Ausprägungen vollständiger Überlappung, partieller Überlappung und keiner Überlappung auf den individuellen Wissenszuwachs der Dyadenmitglieder untersucht. Hierzu wurde das individuelle Systemwissen vor und nach der Gruppendiskussion erhoben und ein Differenzwert bestimmt.

Die systematische Untersuchung von Gruppenleistung (und damit möglicherweise einhergehender Phänomene wie Lernzuwachs) ist ein Kernthema der Sozialpsychologie. Im folgenden Kapitel werden sozialpsychologische theoretische Ansätze und methodische Vorgehensweisen erläutert, die eine angemessene Untersuchung der Gruppenleistung und der möglicherweise im Gruppenkontext auftretenden Lernprozesse bei der dynamischen Systemsteuerung ermöglichen.

### **2.3 Gruppenleistung und Gruppenlernen in der sozialpsychologischen Forschung**

Eine der zentralen Fragen der sozialpsychologischen Forschung ist, ob sich Gruppenprozesse leistungsförderlich oder leistungshemmend auswirken. Um diese Frage zu untersuchen wird die Gruppeninteraktion bei kollektiver Aufgabenbearbeitung in Relation zur Leistung bei vergleichbarer individueller Aufgabenbearbeitung mit den gleichen personalen Ressourcen betrachtet. Hierbei wird die im vorherigen Abschnitt genannte, zweite Fragestellung zur Gruppenleistung fokussiert. Diese zielt darauf ab, ob die interaktive Aufgabenbearbeitung von  $N$  Personen in der Gruppe eine Leistungssteigerung im Vergleich zu der erzielten Leistung der  $N$  Personen ohne gemeinsame Aufgabenbearbeitung bewirkt.

Die Gruppenleistung besteht immer aus einem individualspezifischen Anteil wie der individuellen Leistungsmotivation und individuellen Fertigkeiten und aus einem gruppenspezifischen Anteil wie z.B. gemeinsamer Strategieentwicklung, die über die individuellen Ideen der Gruppenmitglieder hinausgeht (vgl. Drewes, Schultze, & Schulz-Hardt, 2011). Die Leistung, die aus zielführender Kombination der individuellen Leistungen der Gruppenmitglieder ohne Interaktion resultiert, bezeichnet man als Gruppenpotenzial und soll somit den individualspezifischen Anteil abschätzen. Hackman und Morris (1975) nehmen, ebenso wie weitere Forscher auf dem Gebiet der sozialpsychologischen Kleingruppenforschung (z.B. Collins & Guetzkow, 1964; Larson, 2010; Salazar, 1995) an, dass die Gruppenleistung in interagierenden und sozial interdependenten Gruppen unter Umständen die potenzielle Gruppenleistung übertreffen kann. In diesem Fall spricht man von *Prozessgewinnen* (d.h. der gruppenspezifische Anteil ist größer als Null). Prozessgewinne können nach Hackman und Morris (1975) erstens aus interaktionsbedingter Motivationssteigerung der Gruppenmitglieder zur Aufgabenbearbeitung, zweitens aus der interaktiven Entwicklung von Fertigkeiten in Form von Aufgabenbearbeitungsstrategien sowie drittens aus einer Zusammenführung und/oder Neugenerierung von leistungsrelevantem Wissen resultieren. Die sichtbare bzw.

tatsächliche Gruppenleistung übertrifft das Gruppenpotenzial dann, wenn Prozessgewinne vorliegen und gleichzeitig keine oder vergleichsweise geringe Motivationsreduktion, Verschlechterung der individuellen Fertigkeiten oder leistungsmindernde Koordination auftreten. Letztere, durch die Gruppeninteraktion bedingte Leistungsminderungen werden unter dem Begriff *Prozessverluste* subsumiert. Abbildung 1 verdeutlicht diesen Zusammenhang

$$\text{Tatsächliche Gruppenleistung} = \text{Gruppenpotenzial} - \text{Prozessverluste} + \text{Prozessgewinne}$$

Abbildung 1. Formel zur Divergenz zwischen tatsächlicher Gruppenleistung und Gruppenpotenzial in Anlehnung an Hackman und Morris (1975).

Larson (2010) verwendet, ähnlich wie zuvor schon Schweiger und Sandberg (1989) sowie Salazar (1995),<sup>13</sup> den Begriff *Synergie*, um interaktionsbedingte Leistungsgewinne bei der kollektiven Bearbeitung von Aufgaben zu beschreiben:

I define synergy as a gain in performance that is attributable in some way to group interaction. More specifically, a group is said to exhibit synergy when it is able to accomplish collectively something that could not reasonably have been achieved by any simple combination of individual member efforts. Synergy is thus an emergent phenomenon rooted in group interaction. (Larson, 2010, S.4).<sup>14</sup>

Die Begriffe Prozessgewinne und Synergie sind somit weitestgehend synonym zu verwenden. Im Folgenden wird der Übersichtlichkeit halber einheitlich der Begriff Prozessgewinne verwendet, um einen interaktionsbedingten Leistungsvorteil von Gruppen gegenüber dem Gruppenpotenzial zu bezeichnen.

<sup>13</sup> Während Schweiger und Sandberg (1989) dann von Synergie ausgehen, "when a group's performance surpasses that of the capabilities of its group members" (S. 32), hat Salazar (1995) ein erweitertes Begriffsverständnis von Synergie und unterscheidet positive Synergie im Sinne von Larson (2010) und Schweiger und Sandberg (1989) von negativer Synergie für Fälle, in denen die Interaktion von Gruppenmitgliedern zu einer Leistung führt, die unterhalb des Niveaus liegt, das auf Basis der individuellen Fähigkeiten und Fertigkeiten der Gruppenmitglieder zu erwarten wäre (vgl. S. 170). Dieser negative Synergiebegriff entspricht damit den Prozessverlusten nach Hackman und Morris (1975).

<sup>14</sup> Larson differenziert in seiner Konzeption der positiven Synergie darüber hinaus *strong synergy*, die dann vorliegt, wenn die Gruppenleistung die Individualleistung des besten Mitglieds übertrifft, von *weak synergy* im Falle einer Gruppenleistung, die lediglich die Durchschnittsleistung der Gruppenmitglieder (aber nicht die beste Individualleistung) übersteigt.

### 2.3.1 Bestimmung von Prozessgewinnen bei quantitativen Steuerungsurteilen

Um zu untersuchen, ob Prozessgewinne respektive Synergieeffekte aufgrund der Interaktion in der Gruppe auftreten, wird die Gruppenleistung in sozialpsychologischen Studien mit quantitativen Aufgaben häufig mit der post hoc aggregierten Individualleistung einer identischen Anzahl von Individuen (wie Gruppenmitgliedern pro Gruppe) verglichen. Für diese Vergleichsbedingung wird meistens der Begriff *Nominalgruppen* verwendet, in älteren Studien auch die Begriffe *statistized groups* (vgl. Lorge, Fox, Davitz, & Brenner, 1958) oder *Pseudogruppen*.

Eine charakteristische Eigenschaft von Systemsteuerungsurteilen ist, dass die Gruppenmitglieder frei wählen können, in welcher Gewichtung die individuellen Beiträge in das Gruppenprodukt einfließen. Die Gruppe kann beispielsweise das subjektiv beste Einzelurteil eines Gruppenmitglieds als Gruppenurteil auswählen oder die individuellen Vorschläge innerhalb der Gruppe gleichwertig oder auch unterschiedlich stark gewichten, um zu einem Gruppenurteil zu gelangen. Nach der Aufgabentypologie von Steiner (1972) handelt es sich somit um eine *diskretionäre* Aufgabe. Es gibt bei diesem Aufgabentyp keine festgelegte Kombinationsregel der individuellen Steuerungsurteile, um das Gruppenpotenzial zu bestimmen wie bei anderen Aufgabentypen.<sup>15</sup> Eine plausible Potenzialbestimmung besteht darin, bei quantitativen Urteilen generell den Mittelwert der individuellen Urteile der Gruppenmitglieder bzw. einer vergleichbaren Anzahl an Individuen heranzuziehen (vgl. Schultze, Mojzisch, & Schulz-Hardt, 2012). Wenn die Gruppenleistung bei Urteilsaufgaben diesen Mittelwert übersteigt, würden somit Prozessgewinne vorliegen.

Die gemittelte Nominalgruppenleistung bei quantitativen Urteilsaufgaben (wie z.B. Schätz- oder Prognoseaufgaben) berücksichtigt bereits den so genannten *assembly effect bonus* (Collins & Guetzkow, 1964) bzw. den *wisdom of the crowds* Effekt (Surowiecki, 2004). Demnach führt die Aggregation von Informationen bzw. Einzelurteilen in Gruppen

---

<sup>15</sup> Steiner (1972) unterscheidet bei unteilbaren Gruppenaufgaben zwischen den genannten diskretionären Aufgaben, sowie zwischen additiven Aufgaben wie dem Brainstorming, konjunktiven Aufgaben wie dem fehlerfreien Spielen eines Orchesterstückes und disjunktiven Aufgaben wie dem Problemlösen mit einer korrekten Lösungsalternative. Bei additiven Aufgaben setzt sich die Gruppenleistung aus der Summe der Individualleistungen der Gruppenmitglieder zusammen und das Gruppenpotenzial entspricht der Summe der (beim Brainstorming nicht redundanten) individuellen Beiträge der Gruppenmitglieder oder Nominalgruppenmitglieder. Das Gruppenpotenzial bei konjunktiven Aufgaben wird durch die schlechteste Individualleistung eines Gruppenmitglieds bestimmt. Bei disjunktiven Aufgaben entspricht das Gruppenpotenzial der besten individuellen Leistung innerhalb der Gruppe bzw. Nominalgruppe.

häufig dazu, dass ein statistischer Fehlerausgleich stattfindet und die Gruppenurteile besser ausfallen als die Urteile der einzelnen Gruppenmitglieder.<sup>16</sup> Übertrifft die Realgruppenleistung dieses mittlere Nominalgruppenniveau, ist dies wahrscheinlich auf interaktionsbedingte Prozessgewinne zurückzuführen.

Um bei dem speziellen Urteilsaufgabentyp der dynamischen Systemsteuerung mit Sicherheit auf vorliegende Prozessgewinne schließen zu können, genügt es jedoch nicht, die Gruppenleistung mit der post hoc gemittelten Nominalgruppenleistung derselben Anzahl an Gruppenmitgliedern zu vergleichen, die unabhängig voneinander ein dynamisches System durch mehrfache Abgabe von Urteilen gesteuert haben (Bei anderen, statischen Urteilsaufgaben wie quantitativen Prognosen oder Schätzaufgaben wäre dieses Vorgehen des Mitteln *nach* der Aufgabenbearbeitung hingegen angemessen.) Der Grund dafür ist, dass sich der jeweilige Systemzustand während der Steuerungsphase dynamisch in Abhängigkeit von den getroffenen Steuerungsurteilen verändert. Die steuernde Gruppe kann durch aggregierte Steuerungsurteile und individuellen Fehlerausgleich so schon im Verlauf der Steuerungsphase leistungsbegünstigende Systemzustände hervorrufen. Die Individuen produzieren hingegen drei voneinander unabhängige (und wahrscheinlich weniger günstige) Steuerungsverläufe, die somit noch nicht vom Verlauf begleitenden assembly bonus effect profitieren.

Daher führe ich in meiner Arbeit als eine methodische Neuerung die *Nominalgruppenbedingung mit gemitteltem Systemverlauf* ein. Hierbei treffen drei Personen unabhängig voneinander individuelle Steuerungsurteile, die automatisch von einem Computer zu einem Nominalgruppenmittelwert aggregiert werden. Das dynamische System reagiert während des Steuerungsverlaufes auf diese aggregierten Urteile und manifestiert die Nominalgruppenleistung, die aus der kontinuierlichen zielführenden Kombination der Individualurteile während der Steuerung resultiert. Erst wenn die interagierenden Gruppen auch diesem Vergleichsmaßstab überlegen sind, kann man mit Sicherheit von interaktionsbedingten Prozessgewinnen ausgehen.

---

<sup>16</sup> Diese Aussage gilt nur für Fälle, in denen die Gruppe keinen Bias hat (vgl. Larson, 2010), also im konkreten Fall der Systemsteuerung nicht alle Gruppenmitglieder Steuerungsurteile abgeben, die konsistent oberhalb oder unterhalb der optimalen Steuerungsurteile liegen.

### 2.3.2 Differenzierung von Gruppenlernprozessen bei wiederholter kollektiver Aufgabenbearbeitung

Falls Gruppen bei der dynamischen Systemsteuerung Prozessgewinne realisieren, ist eine interessante Frage, worauf dieser interaktionsbasierte Leistungsvorteil beruht. Wenn man davon ausgeht, dass Individuen und Gruppen durch Rahmenbedingungen wie leistungsabhängige Bezahlung gleichermaßen motiviert sind, die bestmögliche Leistung zu erbringen, bieten sich nach Hackman und Morris (1975) noch Gewinne auf der Ebene der individuellen Wissens- und Fertigkeitsbasis und im Bereich der strategischen Aufgabenbearbeitung an. Konkret sind folgende Vorgänge der leistungsförderlichen Gruppeninteraktion denkbar: Erstens könnten die Gruppenmitglieder ihre Annahmen über die Funktionsweise des dynamischen Systems und, falls vorhanden, vorausgehende individuelle Steuerungserfahrungen zusammentragen und ein vollständigeres Systemverständnis erlangen. Zweitens könnte der Austausch unter den Gruppenmitgliedern zu einer gegenseitigen Korrektur fehlerhafter individueller Annahmen über Variablenzusammenhänge und Systemeigenschaften führen. Beide Aspekte würden zu einem verbesserten individuellen Systemverständnis beitragen, was sich dann in einer besseren Steuerungsleistung niederschlagen könnte. Drittens bietet die Gruppeninteraktion die Möglichkeit, effektive individuelle Strategien der Systemsteuerung sowie gute Referenzwerte für Steuerungsurteile zu erkennen und anzuwenden. Die Gruppe könnte somit das Leistungsniveau ihres besten Mitgliedes erreichen. Über die Annahmen von Hackman und Morris (1975) hinaus ist auch denkbar, dass in der Gruppe individuelle Strategievorschläge so kombiniert werden, dass daraus eine bessere Leistung als die des besten Gruppenmitglieds resultiert.

Brodbeck und Greitemeyers (2000a) Kategorisierung von Lernprozessen, die während der kollektiven Aufgabenbearbeitung in der Gruppe auftreten können, ermöglicht für die eben beschriebenen konkreten Lerninhalte eine Zuordnung zu folgenden vier Aufgabentyp unabhängigen Lernprozesskategorien :

- **Individual-to-individual Transfer (I-I Transfer)** beschreibt Übungeffekte, die durch wiederholte Ausführung einer Tätigkeit entstehen. Für diesen Lerneffekt ist zwar keine Interaktion in der Gruppe erforderlich; er kann jedoch auch im Kontext von Gruppenarbeit auftreten. I-I Transfer liegt dann vor, wenn die Einzelleistungen von Individuen bzw. von Nominalgruppen (z.B. bezüglich Systemwissen oder Steuerungserfolg) über die Zeit hinweg besser werden.



- **Group-to-individual Transfer (G-I Transfer)** liegt vor, wenn sich in der Gruppe die individuellen Fertigkeiten der Gruppenmitglieder verbessern. G-I Transfer kann generell empirisch nachgewiesen werden, wenn die Einzelleistungen der Gruppenmitglieder in Gruppen im Laufe der Gruppenphase besser werden als die Einzelleistungen der Personen in den Nominalgruppen (z.B. durch gegenseitige Fehlerkorrektur oder das Lernen von Referenzwerten für gute Steuerungsurteile).
- **Group-to-individual-in-group Transfer (G-IG Transfer)** tritt auf, wenn während der Gruppeninteraktion Fertigkeiten erlernt oder gesteigert werden, die sich ausschließlich im Gruppenkontext auswirken können. Dieser Lerntransfer kann sich positiv auf die Leistung bei beliebiger Gruppenzusammensetzung auswirken. Experimentell kann er nachgewiesen werden, wenn die Mitglieder der Gruppen bessere Einzelleistungen bei einer Aufgabe abgeben als der Mittelwert der Individualeistungen (z.B. wenn die Gruppenmitglieder die Qualität guter Steuerungsvorschläge erkennen und systematisch stärker für die Gruppenurteile gewichten).
- **Group-to-group Transfer (G-G Transfer)** bezieht sich auf die Entwicklung oder Verbesserung von Fertigkeiten, die ausschließlich in einer ganz spezifischen Gruppe wirken können, weshalb man besser von *group-to-individual-in-the-same-group Transfer* sprechen könnte. Dieser Transfertyp kann im Experiment nachgewiesen werden, wenn die Leistung von Gruppen bei langfristiger konstanter Mitgliederzusammensetzung höher ausfällt (z.B. weil die Gruppenurteile am bekannten Expertenmitglied ausgerichtet werden) als die Leistung von Gruppen mit wechselnder Mitgliederbesetzung bei vergleichbar langer individueller Aufgabenerfahrung.

Während *G-IG* und *G-G Transfer* Prozesse beschreiben, die *Koordinationsgewinne durch eine verbesserte Kombination der individuellen Ressourcen innerhalb der Gruppe* ermöglichen, handelt es sich bei *GI Transfer* um einen individuellen Fertigkeitserfolg. Bisher liegt kein abgesicherter Nachweis von Koordinationsgewinnen in der Forschung vor. Es ist ein weiteres Anliegen dieser Arbeit zu überprüfen, ob interagierende Gruppen bei der Steuerung dynamischer Systeme Koordinationsgewinne erzielen und/oder ob die Gruppeninteraktion sich positiv auf die individuellen Fertigkeiten der Gruppenmitglieder auswirkt und *G-I Transfer* resultiert.

### 3. Zielsetzung und Ableitung experimentübergreifender Hypothesen

Mit einer Serie von drei Experimenten soll erstmalig methodisch sauber untersucht werden, ob Gruppen bei der Steuerung dynamischer Systeme interaktionsbedingte Prozessgewinne gegenüber der Leistung einer vergleichbaren Anzahl an Individuen bzw. Nominalgruppen realisieren können. Das zweite Ziel ist es zu klären, ob die überlegene Gruppenleistung durch den wisdom of the crowds Effekt (Surowiecki, 2004) erklärbar ist und entsprechend auf dem Vorteil der statistischen Aggregation der individuellen Urteile basiert, oder ob die *Gruppeninteraktion* zu Leistungssteigerungen führt, die dem Gruppenmittelwert der individuellen Steuerungsurteile (in Nominalgruppen mit gemitteltem Systemverlauf) überlegen sind.

Im vorangehenden Kapitel wurden vielfältige Möglichkeiten genannt, wie eine steuernde Gruppe durch interaktiven Austausch von Systemwissen und strategischen Vorgehensweisen sowie durch die systematische qualitätsbasierte Gewichtung individueller Steuerungsurteile Prozessgewinne gegenüber Nominalgruppen erzielen könnte. Auch die Studie von Wolfe und Chacko (1983) liefert angesichts methodischer Mängel zwar keinen Nachweis, aber einen Hinweis auf mögliche Leistungsvorteile von Gruppen. In Hypothese 1 wird entsprechend das Auftreten von Prozessgewinnen bei der gemeinsamen Systemsteuerung im interagierenden Gruppenkontext vorhergesagt. Wie unter 2.3.1 erläutert, ist für den Nachweis von interaktionsbedingten Prozessgewinnen eine Überlegenheit der Gruppenleistung gegenüber zwei Arten von Nominalgruppen nachzuweisen. Hypothese 1a geht davon aus, dass die Realgruppenleistung der traditionellen Nominalgruppe, in der die Individuelleistungen der Mitglieder post hoc kombiniert werden, überlegen ist. Diese Überlegenheit kann durch den statistischen Aggregationsvorteil bei der Gruppenurteilsbildung und durch darüber hinausgehende interaktionsbedingte Leistungssteigerung entstehen. Mit der Hypothese 1b kann überprüft werden, ob es die Gruppeninteraktion ist, die erforderlich für das Auftreten von Prozessgewinnen ist. Demnach müsste die Gruppenleistung auch den statistischen Aggregationsvorteil von Nominalgruppen mit gemitteltem Systemverlauf übertreffen.

**Hypothese 1:** Gruppeninteraktion führt zu Prozessgewinnen bei der Steuerung eines dynamischen Systems.

**Hypothese 1a:** Gruppeninteraktion führt zu einer besseren Leistung bei der Steuerung eines dynamischen Systems als die Steuerungsleistung von Nominalgruppen mit individuellem Systemverlauf.

**Hypothese 1b:** Gruppeninteraktion führt zu einer besseren Leistung bei der Steuerung eines dynamischen Systems als die Steuerungsleistung von Nominalgruppen mit gemitteltem Systemverlauf.

Da bisher kein gesicherter Nachweis in der Literatur zum Auftreten oder Ausbleiben von Koordinationsgewinnen bei der dynamischen Systemsteuerung vorliegt, wird in dieser Arbeit untersucht, ob generell Koordinationsgewinne auftreten. (Die Differenzierung zwischen G-IG und G-G Transfer wäre dann ein weiterer Schritt falls Koordinationsgewinne nachweisbar wären, und wird an dieser Stelle nicht berücksichtigt.) Um Koordinationsgewinne zu untersuchen, wird die durch die Gruppenurteile erzielte Steuerungsleistung mit der potenziellen Leistung der Gruppe verglichen. Eine plausible Bestimmung des Gruppenpotenzials (sprich der zu erwartenden Gruppenleistung bei zielführender Kombination der Individualbeiträge der Gruppenmitglieder) besteht darin, bei quantitativen Steuerungsurteilen den Mittelwert der individuellen Urteile der Gruppenmitglieder (und die daraus resultierende Steuerungsleistung) heranzuziehen (vgl. Schultze, Mojzisch, & Schulz-Hardt, 2012). Wenn die Gruppenleistung die Mittelwert basierte Leistung übersteigt, würden somit Koordinationsgewinne vorliegen.

In Bezug auf das Vorliegen von Koordinationsgewinnen werden zwei Alternativhypothesen formuliert. Da Personen bei der individuellen Systemsteuerung schon mit der Informationsverarbeitung von zeitlich verzögerten und nonlinearen Systemreaktionen überfordert sind (vgl. Serman, 1994), könnte das zusätzliche Kommunikationsaufkommen in der Gruppe, bei dem vermutlich eine Fülle weiterer Informationen bezüglich möglicher Systemzusammenhänge und möglicher Steuerungsstrategien hinzukommt, die beschränkten Aufmerksamkeits- und Gedächtnisressourcen der Gruppenmitglieder überfordern (vgl. *bounded rationality*; Simon, 1982). Angesichts knapper kognitiver Ressourcen neigen Personen dann dazu auf Urteilsheuristiken (z.B. Reichert & Dörner, 1988) zurückzugreifen, die häufig eine leistungsmindernde Wirkung haben. Dies würde begünstigen, dass interagierende Gruppen keine Koordinationsgewinne erzielen und für die Hypothese 2a sprechen.

**Hypothese 2a:** Die auf Gruppenurteilen basierende Steuerungsleistung ist gleich oder schlechter als die aus dem Mittelwert der individuellen Urteile der Gruppenmitglieder resultierende Steuerungsleistung.

Auf der anderen Seite liefert gerade der Austausch über individuelles Systemwissen und geeignete Steuerungsstrategien eine für das Auftreten von Koordinationsgewinnen günstige Situation. Koordinationsgewinne können dann auftreten, wenn die Gruppe das

Urteil des besten Mitglieds bzw. wiederholt das zu einem Zeitpunkt beste Individualurteil innerhalb der Gruppe als Gruppenurteil auswählt. Voraussetzung ist jedoch, dass die Qualität der individuellen Beiträge zu einem gewissen Grad demonstrierbar ist. Da die Gruppenmitglieder nicht nur über das System diskutieren und abwägen, wie es z.B. bei Prognoseaufgaben der Fall ist, sondern durch gezielte Steuerungseingriffe auch die Angemessenheit von guten Steuerungsurteilen nachweisen könnten, sind Koordinationsgewinne möglich. Eine notwendige, für dynamische Steuerungsaufgaben besondere Voraussetzung der Aufgabedemonstrierbarkeit ist das Verständnis der Gruppenmitglieder für die in der Regel zeitverzögert einsetzende Systemreaktion. Wenn dieses Bewusstsein vorliegt, steigt die Wahrscheinlichkeit für richtige Schlussfolgerungen zu Ursache-Wirkungs-Relationen zwischen Eingabeurteilen und Systemreaktionen und entsprechend könnten die pro Steuerungsdurchgang besten Vorschläge oder das insgesamt beste Gruppenmitglied identifiziert werden. Werden die Gruppenurteile daran ausgerichtet, sind Koordinationsgewinne die über das Gruppenpotenzial hinausgehen, möglich. Alternativ zu einer Auswahl eines Individualurteils kann die Gruppe in der Diskussion auch neue Gruppenurteile generieren, die aus einer anteiligen Gewichtung der Individualurteile resultieren oder auch außerhalb des Wertebereiches der Individualurteile liegen.

**Hypothese 2b:** Die auf Gruppenurteilen basierende Steuerungsleistung übersteigt die aus dem Mittelwert der individuellen Urteile der Gruppenmitglieder resultierende Steuerungsleistung.

Die zweite mögliche Komponente von Prozessgewinnen, individueller Fertigkeitengewinn, der während der Gruppeninteraktion auftritt, wird in Hypothese 3 thematisiert. Aufgrund zahlreicher Nachweise von G-I Transfer bei statischen Gruppenaufgaben wie Regelinduktionsaufgaben (Brodbeck & Greitemeyer, 2000b, Letters-to-Numbers-Tasks (Laughlin, Carey, & Kerr, 2008), mathematischen Problemen (Laughlin & Ellis, 1986) und Distanzschätzaufgaben (Schultze et al., 2012), wird auch das Auftreten von G-I Transfer bei der dynamischen Steuerung für wahrscheinlich gehalten. Die Studie von Meyer und Scholl (2009) liefert zusätzlich einen deskriptiven Hinweis auf individuellen Fertigkeitengewinn nach einer Gruppendiskussion über ein dynamisches System, wenn teilweise heterogenes Ausgangswissen der Gruppenmitglieder gegeben ist. Es ist daher plausibel anzunehmen, dass auch der steuerungsbegleitende Austausch innerhalb von Gruppen zu Fertigkeitengewinnen führt, die über Übungseffekte hinausgehen (vgl. Hypothese 3a). Darüber hinaus kann es leistungssteigernd wirken, wenn für die Urteilsfindung Referenzwerte zur Verfügung stehen (vgl. Laughlin, Bonner, Miner, &

Carnevale, 1999). In Nominalgruppen mit gemitteltem Systemverlauf stellen die automatisch zurückgemeldeten Mittelwerte solche Referenzwerte für die individuellen Steuerungsurteile dar, die ebenfalls einen individuellen Fertigkeitserfolg hervorrufen könnten. Da jedoch weder Systemwissen noch Steuerungsstrategien in diesem Nominalgruppenkontext vermittelt werden, ist es plausibel, dass der G-I Transfer in den Realgruppen größer ausfällt als der Fertigkeitserfolg durch Referenzwertinformation (vgl. Hypothese 3b). Somit werden folgende konkrete Hypothesen zum G-I Transfer formuliert:

**Hypothese 3:** Die Verbesserung der individuellen Steuerungsleistung von Realgruppenmitgliedern fällt größer aus als die Verbesserung der Steuerungsleistung von Nominalgruppenmitgliedern.

**Hypothese 3a:** Die Verbesserung der individuellen Steuerungsleistung von Gruppenmitgliedern in der Realgruppenbedingung fällt größer aus als die Verbesserung der Steuerungsleistung von Nominalgruppenmitgliedern mit individuellem Systemverlauf.

**Hypothese 3b:** Die Verbesserung der individuellen Steuerungsleistung von Gruppenmitgliedern in der Realgruppenbedingung fällt größer aus als die Verbesserung der Steuerungsleistung von Nominalgruppenmitgliedern mit gemitteltem Systemverlauf.

Die Studie von Meyer und Scholl (2009) konnte zeigen, dass eine unstrukturierte Gruppendiskussion innerhalb einer Dyade mit teils wissensheterogenen Probanden zur Verbesserung des dynamischen Systemwissens führte. Daher ist es auch wahrscheinlich, dass Gruppenmitglieder während der gemeinsamen Systemsteuerung durch Austausch von Steuerungsstrategien, Annahmen über die Zusammenhänge der Systemvariablen und Nennung von individuellen Steuerungsurteilen, die als Referenzwerte für eine gute Steuerung dienen, gegenseitig ihre Wissensrepräsentation des Systems erweitern. Auch durch die individuelle Steuerung ist es möglich, das Systemwissen zu erweitern. Der Wissenszuwachs sollte aber geringer ausfallen, als wenn zusätzlich zur Steuerungserfahrung eine Austauschmöglichkeit des individuellen Wissens innerhalb einer Gruppe besteht. Aus diesen Überlegungen resultiert die Hypothese 4 sowie die spezifische Vorhersage der Hypothese 4a, die davon ausgeht, dass Realgruppenmitglieder mehr verbalisierbares Systemwissen erwerben als Nominalgruppenmitglieder. In Bezug auf die Qualität des Systemwissens ist es plausibel anzunehmen, dass Personen während der gemeinsamen Systemsteuerung in der Gruppe richtige Aspekte zur Funktionsweise des Systems oder Variablenzusammenhängen in der Diskussion von den anderen Gruppenmitgliedern erlernen, wenn die Mitglieder ihre Annahmen durch vorangehende Erfahrungen in der individuellen Steuerung begründen. Gleichzeitig können individuell

bestehende, falsche Vorstellungen von Systemzusammenhängen, die in der Gruppe diskutiert werden, korrigiert bzw. verworfen werden. Personen, die ein dynamisches System individuell bzw. ohne die Möglichkeit des Austausches von Systemwissen mit anderen steuerungs erfahrenen Personen steuern, haben keine Möglichkeit, fehlerhafte Vorstellungen zu revidieren oder neue korrekte Sichtweisen zu erlernen. Daraus ergeben sich die Hypothesen 4b und 4c mit spezifischen Teilvorhersagen zur Quantität *korrekter* systembezogener Wissensinhalte und des *Prozentsatzes* korrekter Wissensinhalte in Bezug auf alle individuell vorhandenen Systemannahmen.

**Hypothese 4:** Realgruppenmitglieder erwerben mehr verbalisierbares Wissen über die Systemeigenschaften.

**Hypothese 4a:** Realgruppenmitglieder treffen *mehr Annahmen* über die Systemeigenschaften als Nominalgruppen mit gemitteltem oder individuellem Systemverlauf.

**Hypothese 4b:** Die *Anzahl korrekter Annahmen* über Systemeigenschaften ist bei Realgruppenmitgliedern höher als bei Mitgliedern von Nominalgruppen mit gemitteltem oder individuellem Systemverlauf.

**Hypothese 4c:** Der *Prozentsatz korrekter Annahmen* über Systemeigenschaften ist bei Realgruppenmitgliedern höher als bei Mitgliedern von Nominalgruppen mit gemitteltem oder individuellem Systemverlauf.

## 4. Experiment 1

Experiment 1 ist das erste aus einer Serie von drei Experimenten, in denen erstmals das Auftreten von interaktionsbedingten Prozessgewinnen im Kontext dynamischer Systeme methodisch sauber untersucht werden soll und systematisch zugrundeliegende Lernprozesse in Bezug auf Steuerungsfertigkeiten und Systemwissen untersucht werden. Für das erste Experiment wird ein computersimuliertes, dynamisches Kühlsystemszenario mit relativ geringer Komplexität verwendet. In Experiment 2 wird die Generalisierbarkeit der Befunde in einem vergleichsweise komplexeren dynamischen Wirtschaftsszenario überprüft. Experiment 3 verwendet erneut das Kühlsystemszenario und ermöglicht (neben der Überprüfung der Replizierbarkeit der Ergebnismuster) mit einem erweiterten Design den Nachweis von zeitstabilem G-I Transfer, für den sich in den Experimenten 1 und 2 Hinweise in den Daten fanden.

### 4.1 Zielsetzung und Hypothesen

Ziel des Experimentes 1 ist eine Überprüfung der in Kapitel 3 formulierten Hypothesen im Kontext eines relativ gering komplexen dynamischen Regelkreises, für den die Probanden Steuerungsurteile bezüglich *einer* quantitativen Steuerungsvariable treffen müssen und die Abweichung *einer* quantitativen Outputvariablen von einem optimalen Systemzustand als Leistungsindikator dient. Dieses „einfache“ Szenario, dass aufgrund der Systemeigenschaften trotzdem eine moderate Aufgabenschwierigkeit für die Probanden bietet (vgl. 4.3.1) wurde gewählt, um im ersten Experiment generelle kognitive Überlastung durch ein unüberschaubar großes System (wie beispielsweise Lohhausen mit 2000 Variablen, vgl. Dörner et al., 1983) ausschließen zu können. Bei der Verwendung eines hochgradig komplexen Systems hätte die Gefahr von Bodeneffekten bei der Steuerungsleistung von Individuen sowie von Gruppen bestanden. Ein aufgabenbedingt homogen niedriges individuelles Leistungsniveau hätte das Auftreten von Gruppenlernprozessen erschwert oder sogar unmöglich gemacht.

### 4.2 Stichprobe und Design

An der Untersuchung nahmen 315 Studierende unterschiedlicher Fachrichtungen der Georg-August-Universität Göttingen teil. 179 der Probanden waren weiblich (56.80%). Das Durchschnittsalter der Stichprobe lag bei 23.04 Jahren ( $SD = 3.13$  Jahre). Die Probanden wurden über eine Onlinedatenbank der Abteilung Wirtschafts- und Sozialpsychologie, in der sie sich freiwillig registriert hatten, oder über Aushänge im zentralen Hörsaalgebäude der Universität rekrutiert. Experiment 1 beinhaltet ein 3 x 2

Design mit dem dreistufigen Zwischensubjektfaktor *Bedingung* (Realgruppen- vs. Nominalgruppen- vs. Individualbedingung) und dem zweistufigen Innersubjektfaktor *Steuerungsphase* (Steuerungsphase 1 vs. Steuerungsphase 2). Für die Analyse von Koordinationsgewinnen im Zeitverlauf wird die zweite Steuerungsphase in vier Steuerungsblöcke unterteilt und der Innersubjektfaktor *Trialblock* (1 vs. 2 vs. 3 vs. 4) ergänzt. Für die Detailanalyse von individuellen Fertigungsgewinnen wird außerdem der Innersubjektfaktor *Gruppenmitglied* (bestes vs. mittleres vs. schlechtestes) betrachtet.

## 4.3 Methode

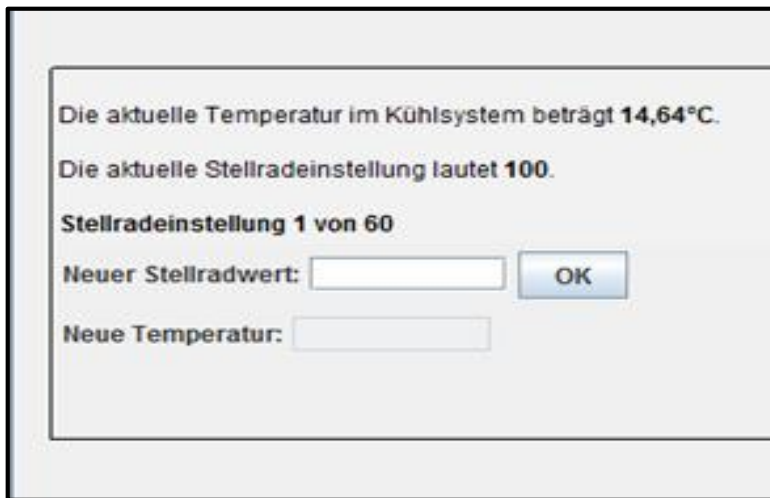
### 4.3.1 Steuerungsaufgabe und Leistungsmaß

In Experiment 1 bestand die Aufgabe in der Steuerung eines dynamischen Systems mit dem Ziel, einen optimalen Systemzustand zu erreichen und beizubehalten. Die Probanden sollten sich vorstellen, sie seien Sicherheitsexperten in einem Atomkraftwerk und nach einem Ausfall des automatischen Kühlsystems dafür verantwortlich, über die Steuerung eines manuellen Ersatzkühlsystems wieder die optimale Kühltemperatur im System herzustellen. Das dynamische Szenario wurde mit der Experimentalsoftware Assembly Lab for Experiments (ALEX; Schlemmer, 2009) programmiert und den Probanden computergestützt dargeboten. Die Software sowie die Experiment-spezifische Datei mit allen Instruktionen, Aufgaben- und Frageitems sind auf Anfrage bei der Autorin erhältlich. Das Aufgabenszenario wurde den Probanden schriftlich am PC präsentiert.

Nach dem Lesen der Instruktion gelangten die Probanden zur Eingabeoberfläche der Systemsteuerung, die in Abbildung 2 dargestellt ist. Zu jedem der 60 Zeittakte einer Steuerungsphase (im Folgenden als Trials bezeichnet) erhielten die Probanden Informationen über die aktuelle Temperatur im Kühlsystem, die auf zwei Dezimalstellen nach dem Komma gerundet in Celsius angegeben wurde. Zu Beginn der Systemsteuerung lag diese Temperatur bei 14.64°C. Für alle nachfolgenden Trials wurde in dieser ersten Zeile die Temperatur angezeigt. In der nächsten Zeile wurde die aktuelle Stellradeinstellung mitgeteilt, die zu Beginn die Ausgangseinstellung 100 betrug. Für alle nachfolgenden Trials wurde an dieser Stelle die vom Probanden vorgenommene Stellradeinstellung des jeweils vorausgehenden Trials gezeigt. Die dritte Zeile informierte über die Trialnummer. Im Beispiel von Abbildung 2 zum ersten Trial wurde entsprechend der Text „Stellradeinstellung 1 von 60“ angezeigt. Im Eingabefeld „Neuer Stellradwert:“ gaben die Probanden dann den für den jeweiligen Trial beabsichtigten Stellradwert an. Durch Klicken auf den „OK“-Button bestätigten sie ihr Urteil und die resultierende „Neue Temperatur“ wurde in der fünften Zeile



ausgegeben. Für die Eingabeurteile war der Wertebereich zwischen 0 und 200 zulässig. Bei Eingabe von negativen Zahlen oder Werten größer als 200 erschien eben dem Eingabefeld die Fehlermeldung „Bitte zulässigen Stellradwert eingeben!“. Die Probanden konnten erst fortfahren, wenn sie eine korrigierte Eingabe im zulässigen Wertebereich vorgenommen hatten.<sup>17</sup> Durch Klicken des „Zur nächsten Einstellung“-Buttons gelangten die Probanden zum Eingabefeld für den nächsten Trial. Hier wurde in der ersten Zeile als „aktuelle Temperatur“, der im vorausgehenden Trial als „Neue Temperatur“ bezeichnete Wert präsentiert.



*Abbildung 2.* Screenshot der auf dem Computerbildschirm angezeigten Ein- und Ausgabeoberfläche zur Systemsteuerung in Experiment 1. Dargestellt ist der Ausgangszustand vor Eingabe der ersten Stellradeinstellung in Steuerungsphase 1 bzw. Steuerungsphase 2.<sup>18</sup>

Nach Eingabe des Stellradwertes in Trial 60 erschien auf dem Monitor die Information, dass die Stromversorgung für das Hauptkühlsystem wieder hergestellt und die manuelle Steuerungsphase damit beendet sei. Die Probanden wurden aufgefordert auf „Weiter“ zu klicken um fortzufahren.

Für ein besseres Verständnis der Systemeigenschaften und Steuerungsanforderungen wird im Folgenden das hinter der sichtbaren Ein- und

---

<sup>17</sup> Bei der Eingabe der Stellradeinstellungen wären auch positive Werte mit Dezimalstellen möglich gewesen. Von dieser Möglichkeit machte aber keiner der Probanden Gebrauch.

<sup>18</sup> Dieser Screenshot stellt aus Lesbarkeitsgründen nur einen Ausschnitt aus der Gesamtansicht für die Probanden dar. Nicht zu sehen sind die Buttons „Zur nächsten Einstellung“ und „Weiter“, die sich rechts außen auf dem Display befanden.

Ausgabefläche liegende Variablensystem erläutert: Dem dynamischen Szenario „Kühlsystem im Atomkraftwerk“ liegen zwei Gleichungen zu Grunde, die in Abbildung 3 dargestellt sind. Dieses Gleichungssystem wurde bereits ohne semantische Einkleidung von Mackinnon und Wearing (1985) und mit der semantischen Einkleidung eines Kühlhauses für Molkereiprodukte von Reichert und Dörner (1988) zur Untersuchung der individuellen Steuerungsleistung verwendet. Für das Experiment 1 wurde der Aufgabenkontext der Kühlsystemsteuerung von einem Warenlager für Molkereiprodukte, die es durch manuelle Steuerung vor dem Verderben zu bewahren galt, in den Kontext der Kühlsystemsteuerung in einem Atomkraftwerk abgewandelt. Hierbei ging es darum, durch Sicherstellen der erforderlichen Kühlung die Gefahr einer Kernschmelze abzuwenden. Dieses neue Szenario der Kühlsystemregulierung sollte die hohe Relevanz einer guten Steuerungsleistung betonen und eine hohe Motivation zur Leistungserbringung sicherstellen.

**Zu Grunde liegende Systemgleichungen.** Die in Abbildung 3 überblicksartig dargestellten semantischen Bezeichnungen der Kühlsystemparameter, sowie die dargestellten Ausgangswerte, Konstanten und Transformationsformeln der resultierenden Werte in einen realistischen Temperaturbereich auf der Celsiusskala, wurden gemäß der Angaben aus der Studie von Reichert und Dörner (1988) übernommen. Demnach beschreiben die Gleichungen (1) und (2) einen Regelkreis, in dem mittels Manipulation einer Stellgröße „stell“ eine Regelgröße „regel“ auf einen bestimmten Sollzustand gebracht werden soll. Die folgenden Erläuterungen der Parameter enthalten zum besseren Verständnis ergänzend die für diese Studie relevanten Kontextbezeichnungen. Gleichung (1) sagt aus, dass die Regelgröße „regel“, hier die zu optimierende Innentemperatur im Kühlsystem, zum Zeitpunkt  $i$  von der vorherigen Innentemperatur „regel <sub>$i-1$</sub> “, der Außentemperatur des Reaktorraumes um das Kühlsystem „stoer“ und der Steuergröße „steuer“ abhängt. Die Steuergröße kann als „die Geschwindigkeit, mit der Kühlmittel durch das System gepumpt wird“ (Reichert & Dörner, S.13), interpretiert werden. Wenn „steuer“ über mehrere Zeittakte den Wert null annimmt, nähert sich die Innentemperatur im Kühlsystem immer weiter der Außentemperatur im Reaktorraum an. Der Parameter „tempo“ kann als Dicke der Außenschicht des Kühlsystems interpretiert werden und beträgt konstant .10. Je kleiner der Wert für „tempo“ ausfällt, desto langsamer passt sich die Innentemperatur im Kühlsystem an die Außentemperatur an. Als letzter Parameter geht die Steuergröße „steuer“ in die Gleichung (1) ein.

(1)	$regel_i = regel_{i-1} + (stoer_i - regel_{i-t}) * tempo - steuer_{i-1}$
(2)	$steuer_i = (regel_{i-v} - stell_i) * regelfaktor$
regel:	Innentemperatur; $regel_3 = 17$ ; $regel_2 = 62.3$ ; $regel_1 = 103.07$ ; $regel_0 = 139.79$ [In °C transformierter Ausgangswert: $(139.79 - 30) / 7.5 = 14.64$ ]
V:	Verzögerungszeit [Konstanter Wert: 3]
$regel_{i-v}$	Innentemperatur 3 Takte vorher [In °C transformierter Ausgangswert: $(62.3 - 30) / 7.5 = 4.31$ ]
stoer:	Außentemperatur [In °C transformierter Ausgangswert: $(170 - 30) / 7.5 = 18.66$ ]
tempo:	Qualität der Isolierung [Konstanter Wert: .10]
stell:	Stellgröße bzw. Stellradgröße [Ausgangswert: 100]
regelfaktor:	Bestimmt Reaktionsstärke des Systems auf Stellradveränderungen [konstanter Wert: .30]
steuer:	Entspricht Geschwindigkeit der Kühlflüssigkeit im System [Ausgangswert: -24.9]

Abbildung 3. Dem dynamischen System in Experiment 1 zugrunde liegende Systemgleichungen und Erläuterungen der Gleichungsparameter. (konstante Ausprägung der Werte bzw. Ausgangswerte zum Zeitpunkt  $i = 1$  in eckigen Klammern).

Gleichung (2) zeigt, dass „steuer“ sich aus der Differenz aus der Regelgröße „regel“ und der Stellgröße „stell“, welche mit dem „regelfaktor“ multiplikativ verknüpft ist, zusammensetzt. Der Regelfaktor bestimmt die Reaktionsstärke des Systems auf Stellradveränderungen und nimmt konstant den Wert .30 an. In die Berechnung der Steuergröße fließt die Verzögerungszeit „v“ ein, deren Ausprägung 3 beträgt. Entsprechend geht der Wert der Innentemperatur „regel“ drei Takte vorher in die Formel ein. Für die Außentemperatur „stoer“ bzw. für die Innentemperatur „regel“ wird mittels der Formel  $(x - 30) / 7.5$ , wobei  $x = stoer$  bzw.  $regel$ , eine Skalentransformation durchgeführt, die analog zur Vorgehensweise in der Studie von Reichert und Dörner (1988) zu Werten führt, die als realistische Temperaturwerte auf der Celsiusskala angesehen werden können. In Abbildung 3 sind für „regel“ und „stoer“ in Klammern die Transformationsberechnungen mit den resultierenden Ausgangswerten der Größen zu Beginn der Systemsteuerung dargestellt. Aus dem noch nicht transformierten Ausgangswert der Außentemperatur „stoer“ von 170 resultiert nach Einsetzen des Wertes in die Transformationsformel mit  $(170 - 30) / 7.5$  beispielsweise eine Ausgangstemperatur von 18.66 Grad Celsius [°C].

Abbildung 4 veranschaulicht das System in einem Blockdiagramm. Die beschriebenen Systemparameter sind als Pfeile dargestellt. Die Kreise visualisieren die Art und Weise des Einflusses eines Parameters auf den anderen. Wenn ein Pfeil bzw. Parameter durch einen grauen Kreissektor in ein Additions-/Subtraktionselement eingeht, wird er verrechnet wie er ist. Weist der Pfeil auf einen entsprechenden schwarzen Sektor, wird der Parameterwert mit -1 multipliziert.

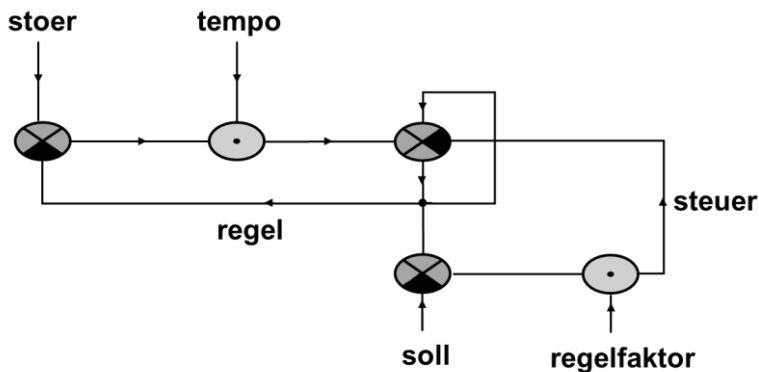


Abbildung 4. Blockdiagramm des dynamischen Systems in Experiment 1.

(● sind Multiplikationselemente, ⊗ sind Subtraktions- bzw. Additionselemente. (übernommen aus Reichert & Dörner, 1988, Abb. 1, S.14).

Die Schwierigkeit der Systemsteuerung bei diesem relativ einfachen Regelkreis mit nur einer Eingabe- und einer Ausgabevariablen ergibt sich erstens aus der Intransparenz des Zusammenhangs zwischen Stell- und Regelgröße, zweitens aus der Zeitverzögerung der Systemreaktion und drittens aus der Nichtlinearität des Systems. Der erste Aspekt der intransparenten Beziehung zwischen der Stellgröße und der zu steuernden Regelgröße macht maßgeblich die Aufgabenstellung der Steuerung aus. Die Probanden erhalten anfangs Informationen zum Wert der Außentemperatur „stoer“ und fortlaufend während der Steuerung Informationen zum vorausgehenden Wert der Innentemperatur „regel“ sowie zu dem nach Eingabe eines Stellwertes „stell“ jeweils aktualisierten Wert der Innentemperatur. Die Probanden müssen die Zusammenhänge zwischen „stell“ und „regel“ durch gezielte Eingriffe und Beobachtung des Systems erschließen, um zielgerichtet das System zu steuern.

Der zweite Aspekt, der wesentlich zur Schwierigkeit der zielgerichteten Steuerung beiträgt, ist die zeitverzögernde Komponente „v“ des Gleichungssystems. Dadurch sind Veränderungen der Stellgröße in Bezug auf die zu regelnde Innentemperatur nicht sofort ersichtlich. Die Probanden müssen die Verzögerungszeit der Systemreaktion auf ihre Eingriffe aus der Beobachtung der Regelgröße über mehrere Zeittakte erschließen. Eine

derartige Verzögerungszeit größer null bewirkt zudem, dass das System gedämpft sinusförmig „schwingt“. Ohne Einschaltung des Ersatzkühlsystems würde sich der Parameter „regel“, also die Innentemperatur des Systems, theoretisch auf den Wert der Außentemperatur „einpendeln“. Das System ist allerdings so programmiert, dass das Ersatzkühlsystem mit dem Ausgangswert der Stellgröße „stell“ auf dem Wert 100 während der gesamten Steuerungsdauer „eingeschaltet“ ist. Bei ausbleibenden Steuerungseingriffen bzw. unter Konstanthaltung des Ausgangswertes der Stellgröße „stell“ auf dem Wert 100 über die 60 Trials einer Steuerungsphase hinweg, würde sich die Innentemperatur um den Wert 11.50 °C einpendeln. In Abbildung 5 ist der entsprechende Temperaturverlauf ohne Steuerungseingriffe dargestellt, dessen sinusförmige Eigenschwingung deutlich zu erkennen ist. Außerdem ist die Range der möglichen maximalen Temperaturwerte bei wiederholter Eingabe der minimalen Stellradeinstellung 0 sowie der maximalen Stellradeinstellung 200 im Zeitverlauf über 60 Steuerungsdurchgänge dargestellt.

Die dritte für das System charakteristische Schwierigkeit bei der Steuerung besteht in der Nichtlinearität des Zusammenhangs zwischen den Veränderungen der Stellradgröße „stell“ und der zu regulierenden Größe „regel“. Erhöhungen bzw. Senkungen des Stellradwertes führen nicht zu einer linearen Transformation des aktuellen Temperaturwertes in einen neuen Temperaturzustand, sondern verursachen eine gedämpft sinusförmige Schwingung der Systemtemperatur über mehrere Steuerungstakte, die sich auf dem neuen höheren bzw. niedrigeren Temperaturniveau einpendelt. Letzteres ist nur der Fall, wenn die Stellradeinstellung über mehrere Zeittakte konstant gehalten wird. Falls nicht, überlagern sich die Auswirkungen der Stellradeinstellungen im Zeitverlauf.

**Leistungsmaß.** Die abhängige Variable „Steuerungsleistung“ wurde über die mittlere prozentuale Abweichung (Mean Absolute Percent Error, kurz MAPE) der erzielten Temperaturwerte von der Zieltemperatur (4°C) über je 59 Trials der Steuerungsphase 1 bzw. Steuerungsphase 2 des Experiments bestimmt. Der erste der 60 Trials einer Steuerungsphase ging nicht in die Auswertung mit ein, da die resultierende Temperatur wegen der Zeitverzögerung der Systemreaktion noch vollständig von den Ausgangseinstellungen determiniert war (und nicht von der ersten Stellradeingabe beeinflusst wurde). Je kleiner die Werte des Abweichungsmaßes MAPE ausfallen, desto besser ist die Steuerungsleistung. Der MAPE ist ein übliches Maß, um die Qualität von Urteilen zu erfassen (z.B. Sniezek & Henry, 1989, 1990; Schultze et al., 2012). Wohingegen in den Studien der genannten Autoren zur quantitativen Schätzleistung direkt die Abweichung der abgegebenen Urteile vom wahren Wert berechnet wird, bezieht sich der

MAPE-Wert in dieser Untersuchung auf die Abweichung der aus den Eingabeurteilen resultierenden leistungskritischen Systemvariablen „Temperatur“.

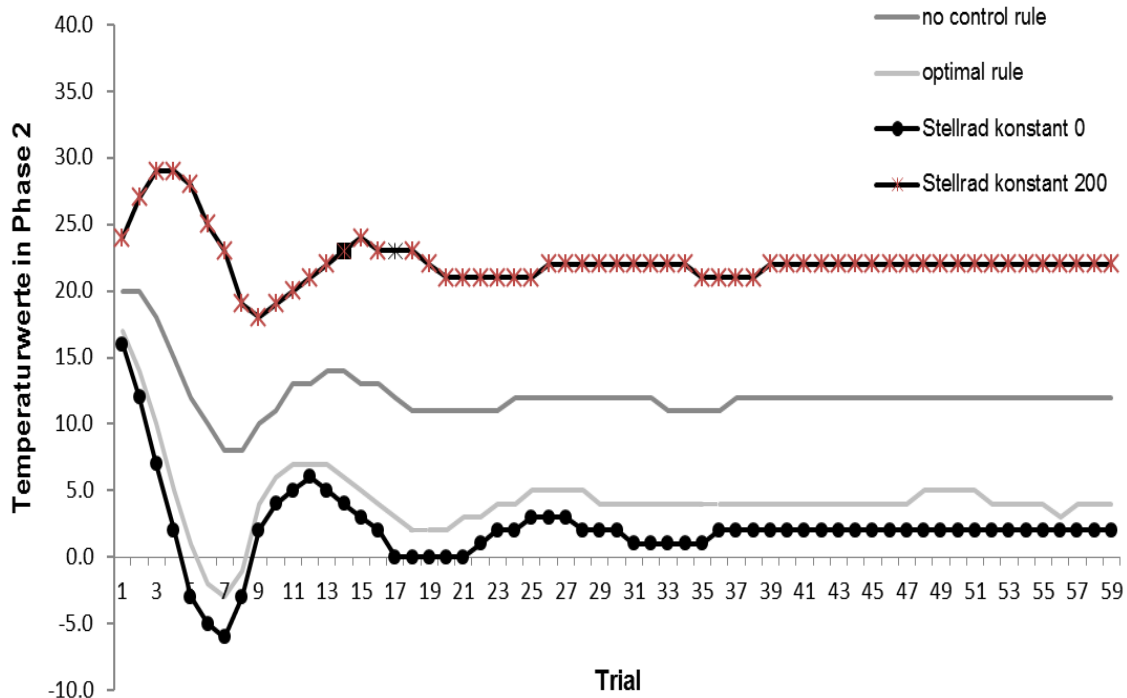


Abbildung 5. Verlauf der Temperaturwerte über die 60 Trials der Steuerungsphase 2 in Experiment 1, die aus der konstanten systembedingt minimalen, der systembedingt maximalen Stellradeinstellung, aus der optimal rule oder aus der no control rule resultierten.

Zur Beurteilung der absoluten Größe des MAPE bieten sich zwei Vergleichsmaßstäbe an, die aus einer optimalen Steuerung bzw. aus dem Ausbleiben der Steuerung, also der eigendynamischen Entwicklung des Systems ohne Eingriffe, resultieren. Die Optimalstrategie zur Minimierung der Temperaturabweichungen von der Zieltemperatur besteht nach Reichert und Dörner (1988) in einer kontinuierlichen Stellradeinstellung von 23. Bei Verfolgung dieser Optimalstrategie (e) resultiert über 60 Steuerungsdurchgänge ein MAPE von .40. Je niedriger der MAPE einer Gruppe ausfällt und je kleiner der Wert ist, um den er diesen optimalen MAPE ( $MAPE_{\text{optimal}}$ ) übersteigt, desto besser ist die erbrachte Steuerungsleistung. Im Experiment von Reichert und Dörner (1988), in dem eine Steuerungsphase 100 Trials umfasste, führte die Eingabe des Stellradwertes 23 zu einer bestmöglichen Annäherung an die Zieltemperatur unter Minimalhaltung der Systemschwingung. Der  $MAPE_{\text{optimal}}$  für eine längere Steuerungsphase von 100 Trials bildet somit die obere Grenze (und damit den geringsten MAPE) des

möglichen Leistungsspektrums bei der Systemsteuerungsaufgabe ab. Die Systemsteuerungsphase im Experiment 1 dieser Arbeit beschränkte sich jedoch auf 60 Steuerungstrials, um eine für die Probanden zumutbare und ökonomische Gesamtdurchführungsdauer des mehrphasigen experimentellen Designs zu gewährleisten. Bei dieser relativ kurzen Steuerungsdauer konnte die überlegene systemstabilisierende Wirkung der Optimalstrategie nicht in vollem Ausmaß im Temperaturverlauf abgebildet werden. Für die Probanden bzw. Gruppen war es zwar schwierig, aber theoretisch möglich, durch von 23 abweichende und über die Trials variierende Stellradeinstellungen einen MAPE zu erzielen, der geringer ausfällt als die konstante optimale Stellradeinstellung 23. Der  $MAPE_{\text{optimal}}$  von .40 markiert somit nicht die absolute Grenze des möglichen Leistungsspektrums der Systemsteuerungsaufgabe in Experiment 1, ist aber ein geeigneter Benchmark für eine gute Steuerungsleistung.<sup>19</sup>

Als zweiter Vergleichsmaßstab für eine sinnvolle Leistungsuntergrenze wird der MAPE herangezogen, der ohne Kontrolleingriffe in die Stellradeinstellung des Ersatzkühlsystems und damit bei Beibehaltung der Ausgangseinstellung des Stellrades auf dem Wert 100 für 60 Zeittakte resultiert. Hierbei spricht man auch von der *no control rule* als Vergleichsmaßstab. Der  $MAPE_{\text{nocontrol100}}$  besteht aus der mittleren absoluten prozentualen Abweichung der Temperaturwerte von der Zieltemperatur bei eingeschaltetem Ersatzkühlsystem mit einem konstanten Stellradwert von 100. In diesem Fall fließt die Steuergröße  $steuer_i$  mit einem Wert  $> 0$  in die Systemgleichung (vgl. Abbildung 3) ein und es resultiert ein  $MAPE_{\text{nocontrol100}}$  von 1.98. Diese untere Leistungsgrenze ermöglicht also Aussagen darüber, ob die von einer Gruppe bzw. von einem Individuum erzielte Steuerungsleistung schlechter oder besser ausfiel als ein Systemverlauf ohne Steuerungseingriffe unter Beibehaltung des Stellradausgangswertes.

---

<sup>19</sup> Theoretisch sind unendlich viele weitere Lösungen mit pro Trial variierenden Stellradeingaben möglich, die in einem vergleichbar niedrigen MAPE-Wert resultieren wie die konstante Stellradeinstellung 23 über alle 60 Trials einer Steuerungsphase. Für einzelne Trials ist es möglich durch von 23 abweichende Stellradeinstellungen eine noch geringere absolute prozentuale Abweichung von der Zieltemperatur zu erzielen. In den Folgetrials würden jedoch bei den meisten Stellradwerten größere Abweichungen aufgrund einer stärkeren Systemschwingung resultieren, was kontraproduktiv für eine Temperaturstabilisierung wäre. In der konkreten Steuerungssituation einer Gruppe ist die für einen Trial optimale Stellradeinstellung immer abhängig von den vorausgehenden Steuerungsurteilen und der daraus resultierenden konkreten Ausprägungen der Systemparameter. Die Optimalstrategie kann demnach als eine gute Steuerungsheuristik für einen fixen Stellradwert bei kontinuierlicher Anwendung angesehen werden. Sie bezieht sich nicht auf die für Einzeltrials optimalen Stellradwerte bei spezifischen, vorausgehenden Systemverläufen.

Der  $MAPE_{nocontrol100}$  stellt nicht die absolut größtmögliche Abweichung der Systemtemperaturen vom Optimalwert dar. Der höchste MAPE (und damit die schlechteste Steuerungsleistung) werden erzielt, wenn in allen 60 Trials die maximal mögliche Stellradeinstellung 200 gewählt wird. Hieraus resultiert ein MAPE von 4.51. Bei einer Wahl der Untergrenze des Wertebereiches von einem Stellradwert von 0 für alle Trials einer Steuerungsphase resultiert ein MAPE von .78, der zwischen dem  $MAPE_{optimal}$  und dem  $MAPE_{nocontrol100}$  liegt. In Abbildung 5 sind die hypothetischen Temperaturverläufe unter der optimal rule und unter der no control rule 100 sowie die tatsächlichen gemittelten Temperaturverläufe in den drei Bedingungen graphisch dargestellt.

Die Temperaturverläufe unter beiden Benchmarks zeigen den für das System typischen, sinusförmig schwingenden Temperaturverlauf. Die Temperaturwerte unter der no control rule 100 fallen (abgesehen von Trial 1) über die gesamte Steuerungsphase höher aus als unter der optimal rule. Das Temperaturniveau unter der no control rule 100 pendelt sich auf einem Niveau um den Wert 11.5 °C ein. Das Temperaturniveau unter der optimal rule nähert sich bei sinusförmiger Schwingung dem Wert der Zieltemperatur von 4 °C an.

**Vortest.** Für die Aufgabe „Kühlsystem im Atomkraftwerk“ wurde ein Vortest mit 30 Probanden durchgeführt um sicherzustellen, dass die Aufgabe zwei notwendige Eignungskriterien (vgl. Schultze et al., 2012) für die Untersuchung von Gruppenlernprozessen erfüllt. Erstens muss bei der Aufgabenbearbeitung ein gewisses Maß an Fähigkeitsstabilität der individuellen Systemsteuerungsleistung vorliegen und resultierende Ergebnisse bei wiederholter Steuerung dürfen nicht nur zufällig variieren. Zweitens muss die individuelle Leistung zwischen den Probanden zu einem gewissen Ausmaß zwischen „schlechter“ und „guter“ Steuerung variieren. Diese Voraussetzungen sind erforderlich, damit eine Heterogenität der individuellen Ausgangsleistung im Gruppenkontext möglich ist und überhaupt Potenzial für interaktionsbedingte Lernprozesse (indem die relativ schlechten Gruppenmitglieder von den relativ guten Mitgliedern lernen) sowie für das Auftreten von Koordinationsgewinnen auf Basis einer stärkeren Gewichtung der besseren Gruppenmitglieder besteht.

Die Auswertung des Vortests, in dem 30 Probanden das System zweimal hintereinander über 60 Zeittakte steuerten, zeigte eine signifikante positive Korrelation zwischen den zwei MAPE-Werten ( $r_s = .56$ ,  $p = .001$ ). Dieses Ergebnis weist auf eine moderate Fähigkeitsstabilität der individuellen Steuerung hin und wurde als hinreichend für den Einsatz der Aufgabe im Hauptexperiment bewertet. Die individuellen MAPE-Werte bei erstmaliger Systemsteuerung variierten zwischen .56 und 2.20, wodurch eine heterogene



Gruppenzusammenstellung gewährleistet werden konnte. Damit erwies sich das Szenario „Kühlsystem im Atomkraftwerk“ als geeignet für die Untersuchung der formulierten Hypothesen.

### 4.3.2 Experimenteller Ablauf

Zu einem Erhebungstermin wurden bis zu 11 Probanden eingeladen, die von zwei oder drei Versuchsleitern betreut wurden. Zu Beginn des Experimentes wurden die Probanden in einem großen Versuchsraum vor separate Laptops gesetzt und von der Versuchsleitung begrüßt und mündlich über den Ablauf des Experimentes informiert. Der gesamte Versuchsablauf an den verwendeten Laptops war standardisiert und ist in Abbildung 6 schematisch dargestellt sowie mit allen erhobenen Items im Originalwortlaut in Anhang A zu finden. Die mündlichen Instruktionen, die die Versuchsleitung erteilte, können in Anhang B nachgelesen werden. Das Experiment wurde den Probanden als eine Untersuchung zum Thema „Systemsteuerungsaufgaben“ vorgestellt. Als Untersuchungsziel wurde Interesse daran genannt, wie Personen individuell oder in der Gruppe diesen bestimmten Aufgabentyp bearbeiten. Anschließend durchliefen die Probanden folgende Phasen. Zunächst machten sie einige demographische Angaben und füllten den NEO-FFI Persönlichkeitsfragebogen (Borkenau & Ostendorf, 1993) aus. Anschließend hatten sie 20 Eingabetakte lang Zeit das System zu explorieren. Darauf folgte eine individuelle Steuerungsphase mit 60 Trials. Basierend auf der Individualleistung teilten die Versuchsleiter die Probanden in leistungsheterogene Gruppen ein und ordneten diese randomisiert einer der drei Bedingungen der Realgruppen-, Nominalgruppen- oder Individualsteuerung zu. Als Kriterium einer leistungsheterogenen Zusammenstellung wurde eine Mindestdifferenz der MAPE-Werte von .10 bei allen paarweisen Vergleichen innerhalb einer Gruppe verwendet. Hierdurch sollte sichergestellt werden, dass die Voraussetzungen für G-I Transfer (bei dem die schlechteren Mitglieder von den besseren lernen) und Koordinationsgewinne aufgrund von G-G-Transfer (die dann auftreten, wenn die Urteile nach Expertise gewichtet in das Gruppenurteil einfließen) in jeder Gruppe gewährleistet sind. Gleichzeitig war das Mindestdifferenzkriterium in Relation zur individuellen Leistungsspanne des Vortests mit MAPE-Werte von .56 bis 2.20 relativ klein gewählt, um nicht künstlich systematisch größere mittlere Leistungsunterschiede innerhalb der Gruppen zu erzeugen, als bei einer zufälligen Gruppenzusammenstellung zu erwarten wären. Den Versuchspersonen wurde zur Zusammenstellung der Gruppen in Phase 2 mitgeteilt, dass es Leistungsunterschiede zwischen den Gruppenmitgliedern geben würde (wobei keine Informationen über die Leistungsreihenfolge oder die konkreten individuellen MAPE-Werte gegeben wurde). Begründet wurde die leistungsheterogene Zusammenstellung damit, dass

eine gleich starke durchschnittliche Leistung für alle Gruppen sichergestellt werden sollte, sodass alle Probanden in den Real- und Nominalgruppen auf Basis der Ausgangsleistung vergleichbare Ausgangsbedingungen für die Leistungserbringung und Bonuszahlung in Phase 2 hatten.

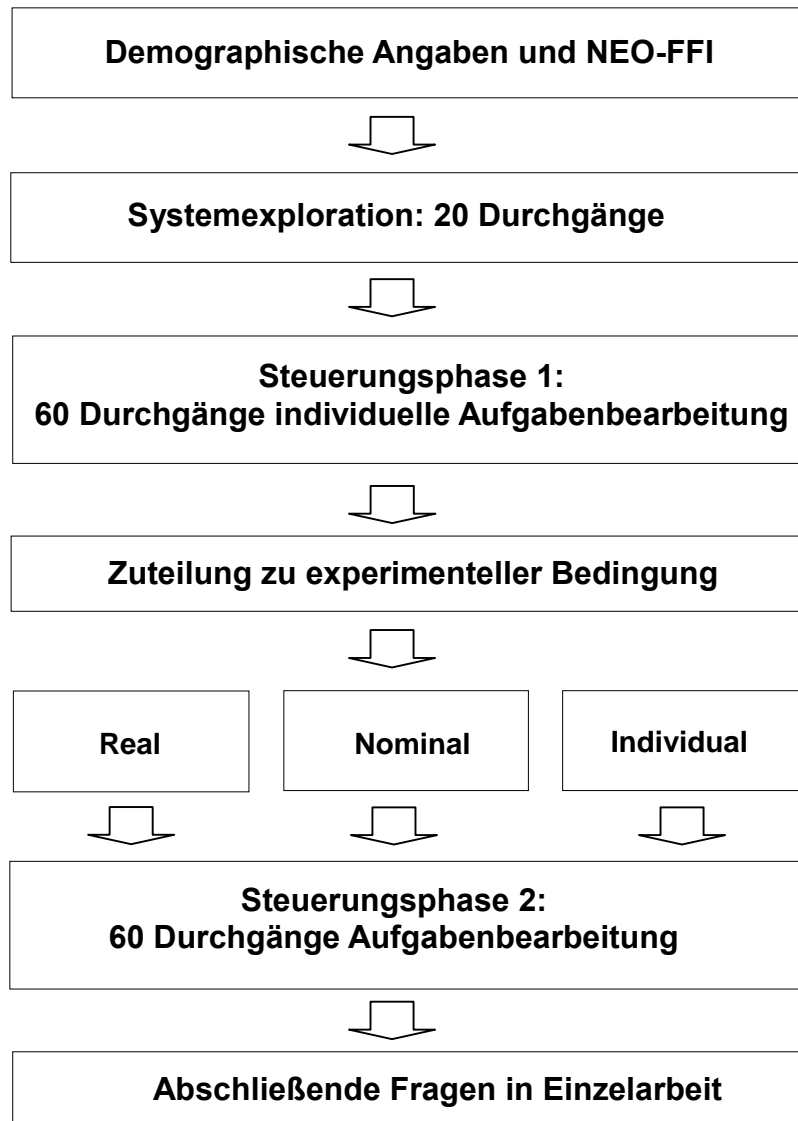


Abbildung 6. Schematische Darstellung des Versuchsablaufes von Experiment 1.

Die Realgruppen wechselten für die Gruppenphase in separate Räume und gaben pro Trial jeweils zuerst ein individuelles Urteil ab und einigten sich dann ohne Vorgaben zum Diskussionsprozess jeweils auf ein Gruppenurteil. Das System reagierte dabei ausschließlich auf die Gruppenurteile. Die Nominalgruppenmitglieder tätigten individuelle Steuerungseingaben und bekamen pro Trial auf ihrem Laptop den Nominalgruppenmittelwert zurückgemeldet. Die Probanden in der Individualbedingung

steuerten das System noch einmal individuell wie in Phase 1. Nach der zweiten Phase erfolgten abschließende Fragen in Einzelarbeit. Nach der Explorationsphase, der Steuerungsphase 1 sowie der Steuerungsphase 2 machten die Probanden Angaben zu ihrer Motivation und ihrem Systemwissen. Der genaue Wortlaut der Items ist in Anhang A nachzulesen und wird im Zuge des Ergebnisberichtes (vgl. Anhang F) beschrieben. Schließlich wurden die Probanden leistungsabhängig entlohnt und es erfolgte ein standardisiertes mündliches Debriefing, dessen Wortlaut in Anhang B nachzulesen ist, sowie der Dank für die Teilnahme und die Verabschiedung.

### **4.3.3 Erfassung des Wissenserwerbs**

Das verbalisierbare Wissen der Probanden über die Eigenschaften und Funktionsweise des Kühlhaussystems wurde jeweils nach der Explorationsphase, sowie nach der ersten und nach der zweiten Steuerungsphase erhoben. Die Probanden sollten ihre konkreten Annahmen über den Zusammenhang von Stellradeinstellung und Kühltemperatur in drei Textfeldern, die mit „Annahme 1“ (bzw. 2 oder 3) betitelt waren, eintragen. Falls sich die Annahmen im Verlauf einer Steuerungsphase geändert hätten, sollten sie sie zeitlich chronologisch ordnen. Bei weniger als 3 Annahmen, wurde ein „X“ in die nicht benötigten Textfelder gesetzt. Die Versuchsleitung ergänzte mündlich die Information, dass bei mehr als drei Annahmen, alle weiteren in das dritte Textfeld geschrieben werden sollten.

Der Inhalt des verbalisierten Wissens wurde von zwei unabhängig voneinander arbeitenden Hilfskräften gemäß eines zuvor erstellten kategorialen Auswertungsschemas (vgl. Anhang C) in einer SPSS-Datendatei kodiert. Hierbei wurde jede Annahme einer von 8 Inhaltskategorien zugeordnet, die vor der Erhebung aus den Systemeigenschaften bzw. zu erwartenden Probandenantworten nach den Erfahrungen von Reichert und Dörner (1988) abgeleitet worden waren. Tabelle 2 zeigt zur Veranschaulichung die Bezeichnungen der 8 Kategorien sowie Beispielnennungen aus dem Vortestdatensatz. Die vier Kategorien mit den Bezeichnungen 0 bis 3 wurden aus den tatsächlichen Eigenschaften des Systems abgeleitet. Sie bezogen sich auf richtige Inhalte. Die vier Kategorien mit den Bezeichnungen 4 bis 7 bezogen sich auf inhaltlich falsche Annahmen und wurden aus den in der Studie von Reichert und Dörner (1988) berichteten Fehlannahmen nach individueller Steuerung des dynamischen Kühlsystems abgeleitet.

Tabelle 2: *Inhaltskategorien zur Kodierung des verbalisierbaren Systemwissens der Probanden in Experiment 1*

Kategorie	Bezeichnung der Inhaltskategorie	Beispielnennungen	Kodierung Richtigkeit
0	optimale Stellrad-einstellung zur Systemstabilisierung richtig	Die Stellradeinstellungen 25-29 erzeugen in etwa eine Temperatur von 4°C	richtig
1	Zusammenhang richtig Hohe (niedrige) Stellradeinstellung führt zu hoher (niedriger) Temperatur	Temperatur sinkt, wenn Stellrad nach unten gedreht wird ODER Große Veränderungen führen zu stärkeren Temperaturänderungen	richtig
2	Zeitverzögerte Reaktion des Systems richtig	Temperatur reagiert erst später ODER System ist träge	richtig
3	„sinusförmige“ Schwingung richtig	Temperaturschwankungen sind auch systembedingt ODER nicht-linear	richtig
4	Zusammenhang falsch	Hohe Stellradeinstellung senkt Temperatur, Zahlen im mittleren Bereich erhöhen die Temperatur	falsch
5	Zeitverzögerte Reaktion des Systems falsch	System reagiert sofort auf Eingaben	falsch
6	Magisches Zahlendenken	1 = hoch, 0 = runter ODER gerade Zahlen sind gut, ungerade Zahlen sind schlecht	falsch
7	Zufall/kein schlüssiges System	willkürlich?	falsch

*Anmerkung:* Für die Inhaltskategorie 1 ist der Wert 23 die korrekte Antwort bei langfristiger Beibehaltung des Wertes. Allerdings kann der als optimal erscheinende Wert je nach produziertem Systemzustand im Einzelfall variieren. Daher wurden Angaben zwischen 18 und 30 als richtig kodiert.

Die Probanden in Experiment 1 dieser Untersuchung trafen pro Eingabetextfeld häufig längere Aussagen, in denen Annahmen zu zwei oder mehr Inhaltskategorien

enthalten waren. Die Kodierer waren angewiesen, dann in SPSS alle relevanten Inhaltskategorien im Annahmefeld mit 1, und alle nicht angesprochenen Inhaltskategorien mit 0 zu kodieren.

Um eine hohe Qualität der Kodierung der Daten in der Hauptuntersuchung sicherzustellen, lasen die Kodierer im Vorfeld einen Informationstext über die Funktionsweise des Systems, wurden mündlich instruiert, erhielten eine schriftliche Kodieranweisung mit den Informationen aus Tabelle 2 und führten schließlich einige Probekodierungen mit Vortestdaten durch, die auf Richtigkeit überprüft wurden.

Nach Abschluss der Kodierung der einzelnen Textfelder wurde pro Proband jeweils ein Summenwert für die Gesamtzahl aller Annahmen, ein Summenwert für die richtigen Annahmen und ein Summenwert für die falschen Annahmen nach der Explorationsphase, nach der Steuerungsphase 1 und nach der Steuerungsphase 2 berechnet. Außerdem wurde der Prozentsatz richtiger Annahmen berechnet, indem die Summe der richtigen Annahmen nach einer Phase durch die entsprechende Gesamtzahl der Annahmen geteilt und anschließend mit 100 multipliziert wurde. Entsprechend wurde bei der Berechnung des Prozentsatzes falscher Annahmen vorgegangen. Diese Berechnungen wurden getrennt für die beiden Datensätze der unabhängigen Kodierer 1 und 2 durchgeführt. Anschließend wurden die berechneten Werte in eine gemeinsame Datei übertragen und die Interraterreliabilität bestimmt. Für jeden Probanden wurde anschließend der Mittelwert der aus den beiden Kodierungen resultierenden Gesamtsumme, der Summe richtiger Annahmen und des Prozentsatzes falscher Angaben berechnet. Dies erfolgte für die Explorationsphase, die Steuerungsphase 1 und die Steuerungsphase 2, sodass 9 gemittelte individuelle Kennwerte als Basis für die weiteren Berechnungen resultierten.<sup>20</sup>

---

<sup>20</sup> Die zwei unabhängigen Kodierer hatten auch die Aufgabe, die offenen Antworten zum individuellen Vorgehen bei der Systemexploration und zum Vorgehen in der Realbedingung und Individualbedingung in Phase 2 zu kodieren. Hierzu waren im Vorfeld der Untersuchung sechs plausible Kategorien für mögliche Vorgehensweisen benannt worden. Die Kodierung ergab, dass der Großteil der Antworten einer Restkategorie mit unspezifischen Angaben der Form „Ja, ich habe meine Vorerfahrungen genutzt“ zuzuordnen war. Somit war keine differenzierte Auswertung des Vorgehens möglich und im Ergebnisteil wird nicht weiter auf diesen Punkt eingegangen. Außerdem wurden die offenen Antworten zum Ziel der Untersuchung gemäß der drei Kategorien 0 = *falsche/keine Zielangabe* oder 1 = *richtig: Vergleich individueller und Gruppenleistung* und/oder 2 = *richtig: Leistung/Lernfortschritt bei der Systemsteuerung* kodiert. Keiner der Probanden musste aufgrund des vermuteten Untersuchungsziels ausgeschlossen werden.

## 4.4 Ergebnisse und Diskussion

Vor der Testung der Hypothesen zu Leistung und Wissenserwerb bei der Steuerung des dynamischen Systems wurde der mögliche Einfluss von Alter, Anzahl der Studiensemester, Studienfach und Geschlecht auf die individuelle Steuerungsleistung überprüft. Als Maß für die Steuerungsleistung diente der MAPE (siehe Kapitel 4.3.1) über die Trials 2 bis 60 einer Steuerungsphase. Das Alter hatte weder einen Einfluss auf die individuelle Leistung in der Steuerungsphase 1, noch auf die Steuerungsleistung in der Individualbedingung in Steuerungsphase 2,  $r(315) = -.04$ ,  $p = .518$  bzw.  $r(105) = .01$ ,  $p = .896$ . Auch die Korrelationen zwischen der Anzahl der bisherigen Studiensemester und der Steuerungsleistung in den Phasen 1 und 2 waren nicht signifikant,  $r(307) = .03$ ,  $p = .633$  bzw.  $r(101) = -.03$ ,  $p = .793$ . Die von den Probanden genannten Studienfächer wurden in die 5 Kategorien Naturwissenschaften, Geisteswissenschaften, Psychologie, Wirtschaftswissenschaften und eine Restkategorie<sup>21</sup> umkodiert. Die Art des Studienfaches hatte keinen signifikanten Einfluss auf die individuelle Leistung der 315 Probanden in der Steuerungsphase 1,  $F(4,310) = 1.15$ ,  $p = .334$ ,  $\eta^2_p = .02$ . Schließlich wurde in einem  $t$ -Test überprüft, ob es Geschlechtsunterschiede bei der individuellen Steuerungsleistung in Phase 1 gab. Männer erzielten einen signifikant niedrigeren MAPE ( $M = 1.14$ ,  $SD = .59$ ) und damit eine bessere Steuerungsleistung als Frauen ( $M = 1.43$ ,  $SD = .58$ ),  $t(313) = 4.38$ ,  $p < .001$ ,  $d = 0.50$ . Auch in der Steuerungsphase 2 fiel die Leistung der 66 weiblichen Probanden in der Individualbedingung ( $M = 1.29$ ,  $SD = .51$ ) signifikant schlechter aus als die Leistung der männlichen ( $M = .86$ ,  $SD = .65$ ),  $t(103) = 3.76$ ,  $p < .001$ ,  $d = 0.74$ . Diese Geschlechtsunterschiede waren für die weiteren Analysen jedoch unproblematisch, da es aufgrund randomisierter Zuteilung der Probanden zu den Bedingungen keine signifikanten Unterschiede der Geschlechtsverteilung in den Bedingungen gab,  $X^2(2, N = 310) = 3.29$ ,  $p = .193$ ,  $\phi = .10$ .

### 4.4.1 Gruppenleistung

**Vergleich mit Leistungsbenchmarks.** Zunächst findet eine Beurteilung des *absoluten Leistungsniveaus* der Realgruppen (sowie der Nominalgruppen mit gemitteltem Systemverlauf und der Nominalgruppen mit individuellem Systemverlauf in der Individualbedingung) statt. Als erster Vergleichsmaßstab für eine sehr gute

---

<sup>21</sup> In die Restkategorie fielen beispielsweise Sportwissenschaften, Geographie oder fehlende Angaben.

Steuerungsleistung wird der  $MAPE_{optimal}$ <sup>22</sup> verwendet, der aus einer konstanten, langfristig optimalen Stellradeinstellung über die Trials einer Steuerungsphase resultiert. Als zweiter Vergleichsmaßstab für eine schlechte Steuerungsleistung wird der  $MAPE_{nocontrol100}$  herangezogen, der bei Ausbleiben von Stellradveränderungen und einer eigendynamischen Entwicklung des Systems resultiert. Anschließend wird untersucht, ob die interagierenden Gruppen Prozessgewinne manifestierten und *in Relation* zu den Vergleichsbedingungen eine signifikant bessere Steuerungsleistung erzielten.

Zur Beurteilung der Steuerungsgüte der in den Bedingungen erzielten MAPE-Werte wurden pro Bedingung je zwei Einstichproben-*t*-Tests gegen den Benchmark-MAPE unter der optimal rule und den Benchmark-MAPE unter der no control 100 rule berechnet. Die herangezogenen Mittelwerte und Standardabweichungen der MAPE-Werte in den drei Bedingungen sind in Tabelle 3 zusammengefasst. Da es möglich ist, beide Vergleichswerte für eine gute bzw. schlechte Steuerungsleistung zu über- und zu unterbieten erfolgte eine .zweiseitige Testung.

Tabelle 3: *Mittelwerte, Standardabweichungen und Konfidenzintervalle der mittleren absoluten prozentualen Abweichung (Mean Absolute Percent Error, MAPE) der erzielten Temperaturen von der Zieltemperatur in den Trials 2 bis 60 der Steuerungsphase 2*

Bedingung	n	M (SD)	95% CI
Real	35	.74 (.42)	[.60, .88]
Nominal	35	.88 (.49)	[.71, 1.04]
Individual	35	1.13 (.35)	[1.01, 1.25]

Anm.: CI = confidence interval (Konfidenzintervall)

Die MAPE-Werte der drei Bedingungen unterschieden sich alle signifikant von dem MAPE-Wert .40 unter der optimal rule, alle  $t(34) > 4.85$ , alle  $p < .001$ , alle  $d > 0.81$ . Die mittlere Steuerungsleistung in allen Bedingungen fiel somit signifikant schlechter aus als die bei Anwendung der Optimalstrategie resultierende Leistung. Deskriptiv kam die Steuerungsleistung der Realgruppen der sehr guten Steuerungsleistung unter Anwendung

<sup>22</sup> Zur Berechnung der Benchmark-Werte wurde analog zur Bestimmung der MAPE-Werte für die Gruppen- und die Individuelleistung die mittlere absolute prozentuale Abweichung der resultierenden Systemtemperaturen von der Zieltemperatur über die Trials 2 bis 60 einer Steuerungsphase bestimmt. Der erste Trial wurde aus der Berechnung ausgeschlossen, da die zugehörige Temperatur von den Ausgangseinstellungen des Systems determiniert war.

der optimal rule näher als die Leistung der Gruppen in der Nominal- und in der Individualbedingung.

Die in den drei Bedingungen erzielten MAPE-Werte in der zweiten Steuerungsphase unterschieden sich ebenfalls alle signifikant von dem Leistungsvergleichswert für eine schlechte Steuerungsleistung unter der no control rule 100 mit einem MAPE von 1.98, alle  $t(34) < -13.44$ , alle  $p < .001$ , alle  $d > 2.24$ . Sowohl die Steuerungsleistung von real interagierenden Gruppen als auch die Steuerungsleistung von Nominalgruppen mit gemitteltem sowie Nominalgruppen mit individuellem Systemverlauf war durchschnittlich besser als der Systemverlauf, der ohne Steuerungseingriffe resultieren würde. Abbildung 7 veranschaulicht den Temperaturverlauf unter den drei Experimentalbedingungen sowie unter den Stellradeinstellungen der optimal rule und der no control rule.

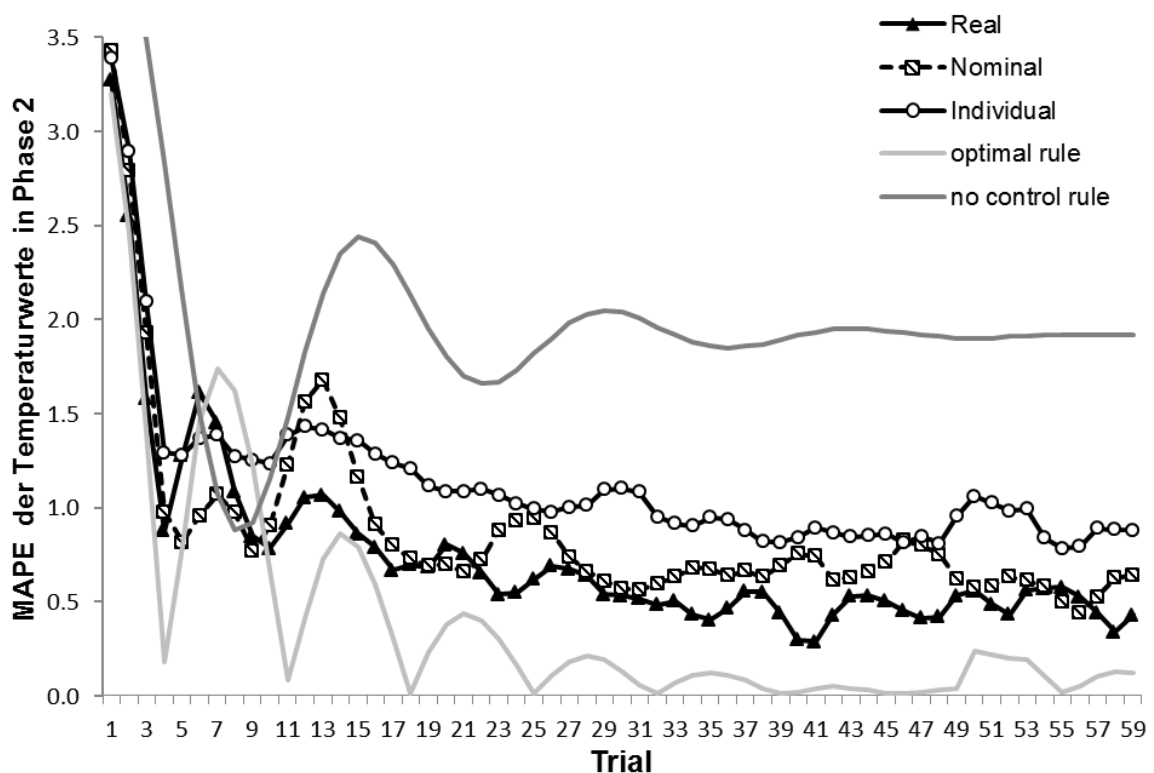


Abbildung 7. Verlauf der MAPE-Werte über die 60 Trials der Steuerungsphase 2, die aus der Steuerung in den drei Bedingungen sowie aus den Benchmarks der optimal rule und der no control rule resultierten.<sup>23</sup>

<sup>23</sup> In Abbildung 7 ist Trial 1, der systembedingt konstante Ausgangswerte unter allen Bedingungen beinhaltet, nicht enthalten.



Zum Abschluss der Benchmark bezogenen Betrachtung der Gruppenleistung wurde überprüft, ob einzelne Gruppen Leistungsergebnisse erzielten, die außerhalb des Leistungsbereiches lagen, der durch die gute Steuerungsleistung mit dem  $MAPE_{optimal}$  und durch die relativ schlechte Steuerungsleistung mit dem  $MAPE_{nocontrol100}$  eingegrenzt wurde. Interessanterweise unterschieden sich die drei Bedingungen deskriptiv bezüglich der Anzahl an Gruppen, die eine vergleichbare oder bessere Steuerungsleistung erbrachten als unter der optimal rule zu erwarten gewesen wäre. In der Realbedingung erzielten 6 der 35 Gruppen einen deskriptiv niedrigeren MAPE als den  $MAPE_{optimal}$ . In der Nominalbedingung konnten nur 3 Gruppen eine bessere Steuerungsleistung erzielen. In der Individualbedingung war keiner der MAPE-Werte der Nominalgruppen mit individuellem Systemverlauf gleich oder kleiner als der  $MAPE_{optimal}$  von .40. Die Häufigkeiten der Gruppen pro Bedingung, die eine schlechtere Steuerungsleistung erbrachten als der Benchmark  $MAPE_{nocontrol100}$ , waren in der Realbedingung und in der Individualbedingung mit jeweils einem Fall identisch. In der Nominalbedingung schnitten zwei Gruppen schlechter ab als der Benchmark für eine relativ schlechte Steuerungsleistung.<sup>24</sup>

**Untersuchung von Prozessgewinnen.** Zur Überprüfung der Hypothese 1 wurde untersucht, ob interagierende Gruppen Prozessgewinne realisierten und eine bessere Leistung bei der Steuerung des dynamischen Systems erbrachten als die Nominalgruppen mit gemitteltem Systemverlauf in der Nominalbedingung und die Nominalgruppen mit individuellem Systemverlauf in der Individualbedingung. Als Leistungsmaß wurde die Differenz aus der Startkompetenz der Gruppen und der Leistung in der Gruppenphase berechnet. Die Startkompetenz bezeichnet die mittlere individuelle Steuerungsleistung der drei Gruppenmitglieder. Sie wurde über den Mittelwert der drei MAPE-Werte der Gruppenmitglieder in der ersten, individuellen Steuerungsphase bestimmt. Die Startkompetenz einer Gruppe stellt einen guten Prädiktor für die Gruppenleistung dar (Schultze et al., 2012). Sie wurde in die Berechnung der Prozessgewinne einbezogen um sicherzustellen, dass auftretende Unterschiede zwischen den Bedingungen in der zweiten Steuerungsphase nicht auf Unterschiede in der individuellen Ausgangsleistung zurückzuführen sind und um die Fehlerstreuung zu reduzieren. Die Leistung in der Gruppenphase wurde mit dem durchschnittlichen MAPE-Wert pro Gruppe über die

---

<sup>24</sup> Auf eine statistische Analyse dieser Häufigkeitsunterschiede zwischen den Bedingungen wurde für beide Benchmarks verzichtet, da der Anteil der erwarteten Häufigkeiten, die kleiner als 5 waren, jeweils 20% überschritt und somit die Voraussetzung zur Durchführung eines  $\chi^2$ -Tests nicht erfüllt war (vgl. Bortz, 1999, S. 170).

Steuerungsdurchgänge der zweiten Steuerungsphase erfasst. Der erste Steuerungsdurchgang (Trial) jeder Phase wurde bei der Berechnung der Steuerungsleistung ausgeschlossen, da die zugehörige Temperatur vollständig durch die Voreinstellung der Systemparameter bestimmt war. Somit gingen in die Berechnung der MAPE-Werte die absoluten prozentualen Abweichungen der erzielten Temperaturwerte von der Zieltemperatur in den Trials 2 bis 60 ein. Die MAPE-Werte in der Realbedingung basierten auf den von den interagierenden Gruppenmitgliedern getroffenen Gruppenurteilen, die einen gemeinsamen Systemverlauf für alle drei Gruppenmitglieder zur Folge hatten. Die MAPE-Werte in der Nominalbedingung resultierten aus dem vom Servercomputer ermittelten Gruppenmittelwert der drei individuellen Steuerungseingaben für jeden Trial. In der Individualbedingung wurde der Mittelwert aus den drei MAPE-Werten für die individuellen Steuerungsverläufe der Gruppenmitglieder bestimmt.<sup>25</sup> Positive Werte des Differenzmaßes zwischen Startkompetenz und gruppenbezogener Leistung in der zweiten Steuerungsphase deuten also auf eine Leistungszunahme in der zweiten Phase hin. Eine Varianzanalyse (ANOVA) mit dem Differenzmaß aus der Startkompetenz der Gruppen und der gruppenbezogenen Leistung in der zweiten Steuerungsphase sowie dem Zwischensubjektfaktor Bedingung (Realbedingung vs. Nominalbedingung vs. Individualbedingung) zeigte einen signifikanten Haupteffekt für den Faktor Bedingung  $F(2,102) = 5.30$ ,  $p = .006$ ,  $\eta^2_p = .09$ . Bonferroni-adjustierte Einzelvergleiche ergaben, dass die auf Interaktion basierende Realgruppenleistung ( $M = .49$ ,  $SD = .37$ ) und die Nominalgruppenleistung ( $M = .46$ ,  $SD = .45$ ) jeweils signifikant besser ausfielen als die Leistung in der Individualbedingung ( $M = .21$ ,  $SD = .37$ ),  $p = .011$ ,  $d = 0.77$  bzw.  $p = .029$ ,  $d = 0.61$ . Die Realgruppenleistung war der Nominalgruppenleistung nur deskriptiv überlegen. Es lag kein signifikanter Unterschied zwischen den beiden Bedingungen vor,  $p = .762$ ,  $d = 0.08$ .

Damit stützen die Daten die Hypothese 1a, die besagt, dass die Systemsteuerungsleistung real interagierender Gruppen der Steuerungsleistung von Nominalgruppen mit individuellem Systemverlauf in der Individualbedingung überlegen ist. Entgegen der Hypothese 1b unterschied sich die Realgruppenleistung jedoch nur deskriptiv

---

<sup>25</sup> Somit entspricht die Individualbedingung einer Nominalgruppenbedingung mit individuellem Systemverlauf. Die in dieser Untersuchung als „Nominalbedingung“ bezeichnete Vergleichsbedingung entspricht hingegen einer Nominalbedingung mit gemitteltem Systemverlauf. Letzteres erfüllt für den Aufgabentyp der dynamischen Systemsteuerungsaufgabe die Forderung von Larson (2010) nach einer zielführenden Kombination von Einzelbeiträgen als Vergleichsmaßstab in der Nominalgruppenbedingung.

und nicht signifikant von der Steuerungsleistung der Nominalgruppen mit gemitteltem Systemverlauf in der Nominalbedingung. Abbildung 8 veranschaulicht die mittleren MAPE-Werte im Verlauf der 60 Trials der Steuerungsphase 2 für die drei Bedingungen. Die hohen MAPE-Werte in den ersten vier Trials sind maßgeblich durch den Ausgangszustand des Systems bedingt. Ab Trial 23 ist das Niveau der mittleren Abweichung vom Temperaturoptimum in der Realbedingung in der Regel niedriger als das Niveau der mittleren Abweichung in der Nominalbedingung. Die mittlere Steuerungsleistung in der Individualbedingung ist in den hinteren 40 Trials der Steuerungsphase 2 durchgängig schlechter als die Nominal- und die Realgruppenleistung.<sup>26</sup>

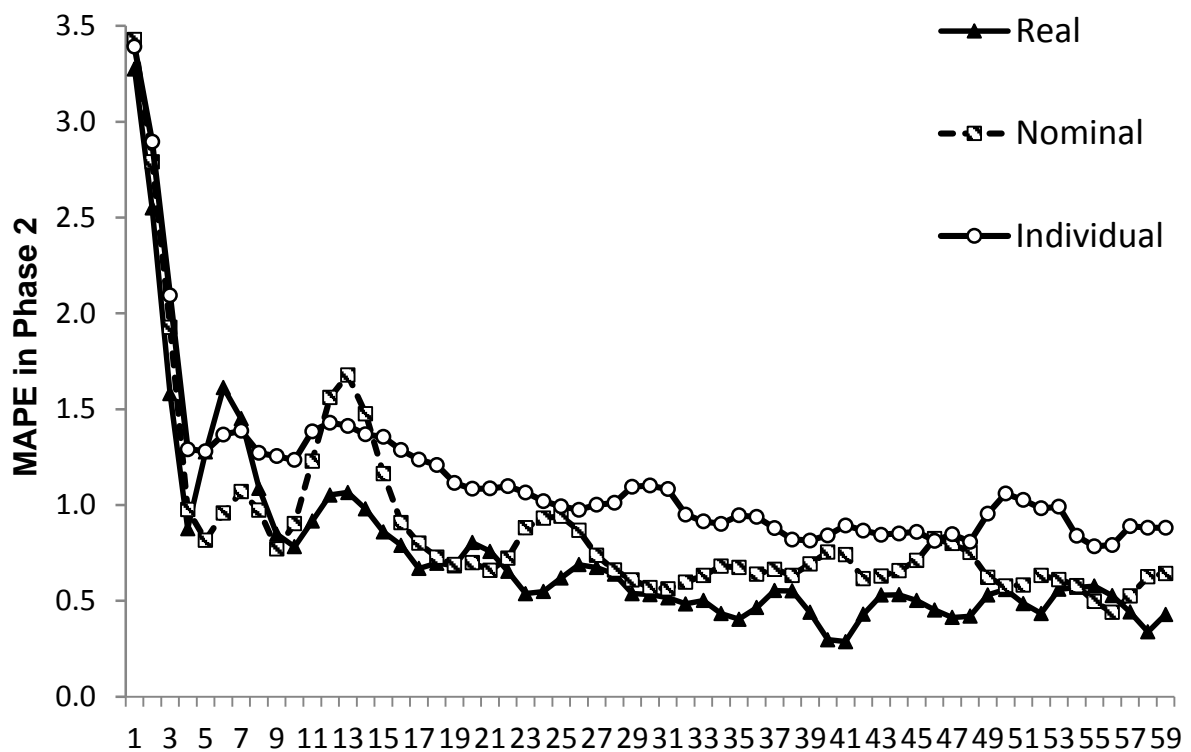


Abbildung 8. MAPE der Temperaturwerte der drei Bedingungen im Zeitverlauf von Phase 2 des Experiments 1.

<sup>26</sup> Ein Vergleich der Steuerungsleistung zwischen den drei Bedingungen für die zweite Hälfte der Phase 2 mit dem MAPE der Trials 31 bis 60 entsprach weitestgehend dem Ergebnismuster der Analyse der Gesamtphase,  $F(2,102) = 6.21$ ,  $p = .003$ ,  $\eta^2_p = .12$ . Die Realgruppenleistung ( $M = .47$ ,  $SD = .52$ ) fiel signifikant besser aus als die Leistung in der Individualbedingung ( $M = .90$ ,  $SD = .45$ ),  $p = .002$ . Die Nominalgruppenleistung ( $M = .64$ ,  $SD = .56$ ) war der Individualleistung hingegen nur noch deskriptiv überlegen mit  $p = .102$ . Real- und Nominalleistung unterschieden sich erneut nicht signifikant voneinander,  $p = .545$ .

#### 4.4.2 Lerneffekte auf Gruppenebene

Die bisherigen Analysen haben gezeigt, dass real interagierende Dreiergruppen eine signifikant bessere Leistung bei der Steuerung des dynamischen Kühlsystemszenarios erbrachten als Nominalgruppen mit individuellem Systemverlauf in der Individualbedingung. Deskriptiv fiel die mittlere Steuerungsleistung über die Gesamtphase 2 leicht besser aus als die durch den statistischen Aggregationsvorteil der Nominalgruppenmitglieder mit gemitteltem Systemverlauf erzielte Abweichung vom Optimum. (Das deskriptiv bessere Leistungsniveau der Realgruppen zeigte sich dabei deutlicher in der hinteren Phasenhälfte.)

In den folgenden Analysen wurde untersucht, ob das höhere Leistungsniveau von Realgruppen auf Koordinationsgewinne zurückzuführen war, die auf einer differentiellen Gewichtung der individuellen Urteile der Gruppenmitglieder basierten. Dazu wurde die Realgruppenleistung mit dem Average Model bzw. dem Gruppenpotenzial verglichen. Das Gruppenpotenzial beschreibt die aus dem Mittelwert der individuellen Urteile resultierende Leistung. In weiteren Analysen wurde überprüft, welches normative, in der Gruppenurteils- und Entscheidungsforschung etablierte Gewichtungmodell die empirische Gruppenleistung am treffendsten beschrieb (vgl. z.B. Sniezek & Henry, 1989; Bonner, Gonzalez, & Sommer, 2004). Hierzu wurde die Passung eines Accuracy Models, das für jeden Trial die Auswahl bzw. vollständige Gewichtung des akkuratesten Individualurteils annimmt, und die Passung eines Best Member Models in zwei Varianten (unter Annahme der Auswahl der Individualurteile des gemäß der Individualleistung in Phase 1 bzw. in Phase 2 besten Gruppenmitglieds) bestimmt. Je geringer der Unterschied zwischen der von einem Modell vorhergesagten Leistung und der empirischen Gruppenleistung ausfällt, desto besser beschreibt das Modell das Gruppenleistungsniveau und die möglicherweise damit einhergehende Gewichtungsstrategie zur Festlegung der Gruppenurteile.<sup>27</sup>

**Gruppenleistung vs. Average Model.** In einer ersten Analyse wurde die auf Basis der Realgruppenurteile erfasste Steuerungsleistung mit der Steuerungsleistung verglichen, die daraus resultiert wäre, wenn pro Trial der Mittelwert aus den drei individuellen Urteilen

---

<sup>27</sup> Auch bei sehr guter Passung eines normativen Modells auf die Gruppenleistungsdaten bedeutet dies nicht zwingend, dass das Gewichtungsschema in der Gruppe umgesetzt wurde. Die Passung kann lediglich mit Sicherheit aufzeigen, dass die Gruppenleistung vergleichbar zur entsprechenden normativen Leistung ist. Ein bestimmter MAPE der Steuerungsleistung unter dem Average Model über alle 60 Trials der Steuerung könnte aber beispielsweise von der Gruppe auch dadurch erzielt werden, dass sie in den Trials alternierend die Individualvorschläge aller drei Gruppenmitglieder als Gruppensteuerungsurteile auswählt.

der Gruppenmitglieder (vor Diskussion einer Gruppenlösung) als Stellradeinstellung realisiert worden wäre. Die theoretisch aus dem Average Model bzw. dem Gruppenpotenzial resultierenden Temperaturwerte wurden nachträglich mit der Statistiksoftware SPSS berechnet. Anschließend wurden pro Gruppe die absoluten prozentualen Abweichungen der gruppenurteils- bzw. der gruppenpotenzialbasierten Temperaturen von der optimalen Zieltemperatur für jeden Trial bestimmt. Da es plausibel erscheint, dass Gruppen erst nach einiger Zeit der gemeinsamen Aufgabenbearbeitung lernen, Individualurteile nach der Mitgliederexpertise oder der Urteilsakkuratheit stärker zu gewichten, wurde der MAPE jeweils für das Gruppenpotenzial und die Realgruppenleistung im Zeitverlauf der Steuerungsphase 2 erhoben. Hierzu wurden die 60 Trials in 4 Blöcke à 15 Trials aufgeteilt, über die jeweils MAPE-Werte berechnet wurden. Der erste Trial wurde wie bisher aus den Analysen ausgeschlossen, da die entsprechende Systemtemperatur vollständig durch den Ausgangszustand der Systemeinstellungen determiniert war. Daher beinhaltete der erste Trialblock nur 14 Trials. Es wurde entsprechend eine MANOVA mit einem 2 (Leistungsmaß: Realgruppenleistung vs. Gruppenpotenzial) x 4 (Trialblock: 1 bis 4) Design durchgeführt. Der Zeitfaktor Trialblock war signifikant,  $F(1.53, 51.85)^{28} = 49.07$ ,  $p < .001$ ,  $\eta^2_p = .59$ . Der von Block 1 zu Block 2 stark und zwischen den folgenden Trialblöcken weniger stark abfallende Verlauf der MAPE-Werte wurde durch einen linearen sowie einen quadratischen Innersubjektkontrast beschrieben,  $F(1, 34) = 53.22$ ,  $p < .001$ ,  $\eta^2_p = .61$  bzw.  $F(1, 34) = 62.60$ ,  $p < .001$ ,  $\eta^2_p = .65$ . Der für die Testung der Hypothese 2 relevante Faktor Leistungsmaß war nicht signifikant,  $F(1, 34) = 1.458$ ,  $p = .236$ ,  $\eta^2_p = .041$ . Die Realgruppenleistung in allen vier Blöcken der Steuerungsphase 2 war bis zur dritten Nachkommastelle der MAPE-Werte identisch mit der erwarteten Leistung bei Mittelung der individuellen Urteile (Trialblock 1:  $M = 1.38$ ,  $SD = .47$ , Trialblock 2:  $M = .68$ ,  $SD = .52$ , Trialblock 3:  $M = .46$ ,  $SD = .51$ , Trialblock 4:  $M = .48$ ,  $SD = .58$ ).<sup>29</sup> Wie bei nahezu identischem Verlauf der Leistungsmaße zu erwarten war, lag keine signifikante Interaktion

---

<sup>28</sup> Bei dieser Analyse zeigten signifikanter Mauchly-Tests die Verletzung der Sphäritätsannahme der zugrundeliegenden Daten für das Leistungsmaß sowie für die Interaktion Trialblock x Leistungsmaß an, beide  $p < .001$ . Daher wurde für die  $F$ -Werte eine Korrektur der Freiheitsgrade nach Greenhouse-Geisser angewendet. Dieses Vorgehen erfolgte konsistent bei allen folgenden Analysen mit verletzter Sphärität.

<sup>29</sup> Ein Wilcoxon-Vorzeichen-Rang-Test für verbundene Stichproben mit der Summe der absoluten prozentualen Abweichungen in der Steuerungsphase 2 als abhängige Variable zeigte, dass Gruppen eine signifikant schlechtere Leistung erbrachten als es das Gruppenpotenzial ermöglicht hätte,  $z = 3.55$ ,  $p < .001$ , zweiseitige Testung: Demnach erzielten 32 der 35 Realgruppen (minimal) größere Abweichungen in Bezug auf die Zieltemperatur als das Gruppenpotenzial bzw. Average Model.

zwischen Trialblock und Leistungsmaß vor,  $F(1.48, 50.24) = .150$ ,  $p = .80$ ,  $\eta^2_p = .00$ . Als zentrales Ergebnis der ANOVA konnte festgestellt werden, dass die Gruppeninteraktion nicht zu Koordinationsgewinnen bei der Steuerung des Kühlsystemszenarios führte. Daraus folgt die Annahme der Alternativhypothese 2a, die davon ausgeht, dass die auf Gruppenurteilen basierende Steuerungsleistung gleich oder schlechter als die aus dem Mittelwert der individuellen Urteile der Gruppenmitglieder resultierende Steuerungsleistung ausfällt. Real interagierende Gruppen realisierten keine Koordinationsgewinne bei der dynamischen Systemsteuerung, die über das Gruppenpotenzial hinausgehen.

Die Realgruppen schienen bei der Einigung auf ein Gruppenurteil eine Mittelungsstrategie zu verfolgen und die individuellen Urteile nicht systematisch nach Akkuratheit, Expertise oder Extraversion zu gewichten. Um diese Schlussfolgerung noch weiter abzusichern, wurde die empirische Gruppenleistung im Folgenden mit der zu erwartenden Steuerungsleistung unter dem Accuracy Model, unter den beiden Expertisemodellen des Initial Best Member und des True Best Member sowie unter einem Extraversionsmodell verglichen.

**Passung des Average Models und weiterer Vergleichsmodelle: Root Mean Squared Difference (RMSD):** Als Kennzahl für die Passung des Average Models und für einen explorativen Vergleich der relativen Passung mit den genannten weiteren normativen Modellen der Gruppenleistung wurde der RMSD bestimmt. Hierzu wurde für jedes Modell die Wurzel aus den mittleren quadrierten Abweichungen der unter den Vergleichsmodellen resultierenden Temperaturen von den unter den Realgruppenurteilen resultierenden Temperaturen der Trials 2 bis 60 der Steuerungsphase 2 berechnet. Je kleiner die resultierende Kennzahl Root Mean Squared Difference (RMSD) ausfällt, desto besser beschreibt ein Vergleichsmodell die aus der Realgruppensteuerung resultierende Leistung. Wie oben beschrieben resultierte die Temperatur unter dem Average Model aus den post hoc pro Trial gemittelten Individualurteilen der Realgruppenmitglieder. Zur Berechnung der Steuerungsleistung des Accuracy Models wurde für jeden der Trials 2 bis 60 der Steuerungsphase 2 das individuelle Urteil desjenigen der drei Realgruppenmitglieder ausgewählt, das den Temperaturwert mit der geringsten Abweichung zur Zieltemperatur zur Folge hatte. In vielen Studien, die normative Modelle mit der Gruppenleistung vergleichen, wird die Leistung des durchschnittlich besten Gruppenmitglieds in der individuellen Steuerungsphase, die der Gruppenphase im I-G Design vorausgeht, als Vergleichskriterium herangezogen. Entsprechend wurde die Realgruppenleistung in Steuerungsphase 2 mit der Initialleistung des pro Gruppe im Durchschnitt besten Mitglieds in der Steuerungsphase 1 verglichen. Dieses Expertisemodell wird hier als *Initial Best*

*Member Model* (vgl. Schultze, 2007) bezeichnet um es von dem anschließend berechneten *Best Member True Model* mit Bezug auf die Leistung in der Steuerungsphase 2 abzugrenzen. In dieser Untersuchung wird abweichend von bisherigen normativen Vergleichsstudien im I-G- Design ein Best Member Model gemäß der besten Gruppenmitgliedsleistung in der Gruppenphase als methodisch angemessen angesehen.<sup>30</sup> Es ist plausibel anzunehmen, dass es innerhalb der Gruppen Änderungen in der Leistungsreihe von Phase 1 zu Phase 2 gab. Das traditionelle (bzw. hier Initial Best Member) Model würde in solchen Fällen dem falschen Gruppenmitglied die Expertise zuordnen und die Leistung des besten Gruppenmitglieds unterschätzen. Eine dem traditionellen Best Member Model gleichwertige oder überlegene Gruppenleistung könnte somit fälschlicherweise als Koordinationsgewinn und Beleg für G-G Transfer interpretiert werden. Daher wurde als weiteres normatives Vergleichsmodell das *True Best Member Model* berechnet, bei dem als Grundlage für den RMSD die aus den Individualurteilen des in Phase 2 besten Gruppenmitglieds resultierenden Temperaturabweichungen von der Optimaltemperatur in den RMSD-Wert eingingen

Die RMSD-Werte betragen für das Average Model bzw. Gruppenpotenzial .03, für das Accuracy Model .80, für das Initial Best Member Model .64 und für das Best Member Model .82. In separaten Wilcoxon-Tests zeigte sich, dass der RMSD des Average Models signifikant niedriger ausfiel als die RMSD-Werte des Accuracy Models, des Initial Best Member Models und des Best Member Models, alle  $z = 5.16$ , alle  $p < .001$ . Das Average Model bzw. die aus dem Gruppenpotenzial resultierenden Temperaturwerte hatten die beste Passung, um die aus den Realgruppenurteilen resultierenden Steuerungsergebnisse zu beschreiben.

Mit einem weiteren Modell wurde betrachtet, ob die Extraversion der Gruppenmitglieder in den Realgruppen eine Auswirkung auf die Gewichtung hatte. Zahlreiche empirische Forschungsbefunde deuten darauf hin, dass in ad hoc zusammengestellten Gruppen ohne Vorerfahrung bezüglich der relativen Mitgliederkompetenz der Einfluss extravertierterer Gruppenmitglieder auf das Gruppenergebnis größer ausfällt, wenn die Gruppe Aufgaben mit geringer bzw. eingeschränkter Demonstrierbarkeit der richtigen Lösung bearbeitet (vgl. Bonner & Sillito,

---

<sup>30</sup> Bei der verwendeten dynamischen Systemsteuerungsaufgabe traten, wie noch in Abschnitt 4.4.3 und vertiefend im Anhang D dargestellt wird, individuelle Fertigkeitsergebnisse bei wiederholter Aufgabenbearbeitung auf. Die initial besten Gruppenmitglieder der Real- und der Nominalbedingung verschlechterten sich deskriptiv, während die mittleren und schlechtesten Gruppenmitglieder ihre Leistung in Phase 2 signifikant steigerten.

2011; Bonner, Sillito, & Baumann, 2007). Im Folgenden wurde die tatsächliche Gewichtung nach Extraversion anhand der Passung eines *Extroversion Models auf die Gruppenleistungsdaten überprüft. Dieses Modell geht von einer vollständigen Gewichtung der Individualurteile des extravertiertesten Gruppenmitglieds aus*. Der RMSD des Extraversion Models betrug .65 ( $SD = .54$ ). Die Passung des Extraversion Models wurde ebenfalls mit dem RMSD des Average Models von .03 verglichen. Ein Wilcoxon-Test zeigte, dass der RMSD des Average Models signifikant niedriger ausfiel als der des Extroversion Models,  $Z = 5.16$ ,  $p < .001$ . Für alle 35 Realgruppen passte das Average Model besser auf die Gruppenleistungsdaten. Anhang E beinhaltet die Darstellung bisheriger empirischer Befunde zur Rolle von Extraversion in Gruppendiskussionen und weitere Analysen zum Zusammenhang von Extraversion mit der relativen Kompetenzeinschätzung und der subjektiven Gewichtung der Gruppenmitglieder in Experiment 1.

#### 4.4.3 Lerneffekte auf individueller Ebene

In weiteren Analysen wurde das Auftreten von I-I Transfer (Übungseffekten) in der Individualbedingung und von individueller Leistungsverbesserung in der Realbedingung und in der Nominalbedingung untersucht. Hierzu wurde für jeden Probanden die Differenz aus dem individuellen MAPE-Wert der ersten Steuerungsphase und dem aus den individuellen Steuerungsurteilen resultierenden MAPE-Wert der zweiten Steuerungsphase berechnet.<sup>31</sup> Bei diesem Differenzmaß deuten positive Werte auf eine individuelle Leistungssteigerung von der ersten zur zweiten Steuerungsphase hin. Da die individuellen Steuerungsurteile in der Realbedingung und in der Nominalbedingung alternierend zu den Gruppenurteilen bzw. den vom Serverrechner zurückgemeldeten Gruppenmittelwerten abgegeben wurden und der jeweilige Systemzustand durch die gruppenbasierten Urteile hervorgerufen wurde, waren die Urteile innerhalb einer Gruppe nicht unabhängig voneinander. Entsprechend wurden die Lerneffekte auf Gruppenebene ausgewertet, indem

---

<sup>31</sup> In der ersten Steuerungsphase bestimmten die individuellen Steuerungsurteile in allen drei Bedingungen den tatsächlichen Temperaturverlauf des Systems. Ebenso resultierte der evidente Temperaturverlauf der zweiten Steuerungsphase in der Individualbedingung aus den individuellen Stellradeinstellungen. In der Realbedingung und in der Nominalbedingung wurden die individuellen Steuerungsurteile in der zweiten Steuerungsphase hingegen nur systemintern abgespeichert, ohne dass die daraus resultierenden Temperaturen den Probanden präsentiert wurden. Die aus den individuellen Urteilen in der Real- und in der Nominalbedingung theoretisch resultierenden Temperaturen wurden nachträglich mit SPSS berechnet. Wie schon im Methodenteil beschrieben, resultierte der Verlauf der Innentemperatur des Kühlsystems in der zweiten Steuerungsphase in der Realbedingung aus den Gruppenurteilen zur Stellradeinstellung und in der Nominalbedingung aus dem vom Server berechneten Mittelwert aus den drei individuellen Steuerungsurteilen der Gruppenmitglieder.



der Faktor Gruppenmitglied als Innersubjektfaktor in die Analyse aufgenommen wurde. Um das Ausmaß individueller Lerneffekte zwischen den experimentellen Bedingungen zu vergleichen wurde entsprechend eine 3 (Bedingung: Realbedingung vs. Nominalbedingung vs. Individualbedingung) x 3 (Gruppenmitglied: bestes vs. mittleres vs. schlechtestes) Kovarianzanalyse (ANCOVA) mit der Bedingung als Zwischensubjektfaktor, dem Gruppenmitglied als Innersubjektfaktor und der Startkompetenz der Gruppen als Kovariate berechnet. Der Status der Gruppenmitglieder als bestes, mittleres und schlechtestes Gruppenmitglied resultierte aus der relativen individuellen Leistung in der ersten Steuerungsphase. Die Startkompetenz der Gruppe stellte das mittlere individuelle Leistungsniveau der Gruppe in der ersten Steuerungsphase dar. Sie hat Einfluss darauf, wie groß das individuelle Verbesserungspotenzial der Gruppenmitglieder ist und wurde als Kovariate in die Analyse einbezogen, um die Fehlervarianz zu reduzieren. In der ANCOVA zeigte sich ein signifikanter Haupteffekt für die Kovariate Startkompetenz,  $F(1, 101) = 7.86$ ,  $p < .01$ ,  $\eta^2_p = .07$ . Dies ist darauf zurückzuführen, dass die Gruppenmitglieder in Gruppen mit einer hohen Startkompetenz, respektive einem schlechteren Ausgangsniveau der mittleren Individualleistung, höhere individuelle Fertigkeitsszuwächse erzielten,  $r(105) = .22$ ,  $p = .027$ . Der Haupteffekt für den Faktor Bedingung war ebenfalls signifikant,  $F(2, 101) = 6.16$ ,  $p = .003$ ,  $\eta^2_p = .11$ . Post-hoc Einzel-t-Tests gegen Null zeigten, dass sich die individuellen MAPE-Werte in allen Bedingungen signifikant verkleinerten. Die Differenz in der Realbedingung ( $M = .48$ ,  $SD = .37$ ),  $t(34) = 7.81$ ,  $p < .001$ ,  $d = 1.32$ , fiel dabei deskriptiv größer aus als die Verbesserung in der Nominalbedingung ( $M = .41$ ,  $SD = .44$ ),  $t(34) = 5.55$ ,  $p < .001$ ,  $d = 0.94$ . In der Individualbedingung war die gegen Null getestete Verbesserung der Steuerungsleistung im Vergleich zu den anderen zwei Bedingungen am geringsten ( $M = .21$ ,  $SD = .37$ ),  $t(34) = 3.31$ ,  $p < .01$ ,  $d = 0.56$ . Post-hoc Tests mit Bonferroni-Korrektur zeigten, dass die individuelle Leistungssteigerung in der Realbedingung signifikant größer ausfiel als die Übungeffekte in der Individualbedingung,  $p = .002$ ,  $d = 0.73$ . Die individuelle Leistungssteigerung in der Nominalbedingung war deskriptiv ebenfalls größer als in der Individualbedingung. Der Unterschied war mit  $p = .076$  und einem Effekt von  $d = .49$  jedoch knapp nicht signifikant. Die Verbesserung der individuellen Steuerungsleistung in der Realbedingung war deskriptiv größer als in der Nominalbedingung, der Unterschied wurde jedoch nicht signifikant,  $p = .691$ ,  $d = 0.18$ . Die bisherigen Ergebnisse belegen, dass ein I-I Transfer bei wiederholter Steuerung der dynamischen Systemsteuerungsaufgabe im Kühlsystemszenario auftrat. Die signifikante Steigerung der Steuerungsleistung, basierend auf den individuellen Urteilen in der Real- und in der Nominalbedingung, die signifikant bzw. deskriptiv über reine Übungeffekte hinausgingen, könnte darauf hindeuten, dass durch den Kontext der Gruppenmittelwertbildung bzw. durch die Interaktion in der Gruppe ein

Lerntransfer von der Gruppe auf das Individuum stattfand. Die individuelle Leistungssteigerung kann jedoch aus methodischen Gründen nicht eindeutig als G-I Transfer interpretiert werden, was am Ende dieses Abschnitts genauer erläutert wird. Auf eine Ergebnisinterpretation in Bezug auf Hypothese 3 wird daher an dieser Stelle verzichtet.

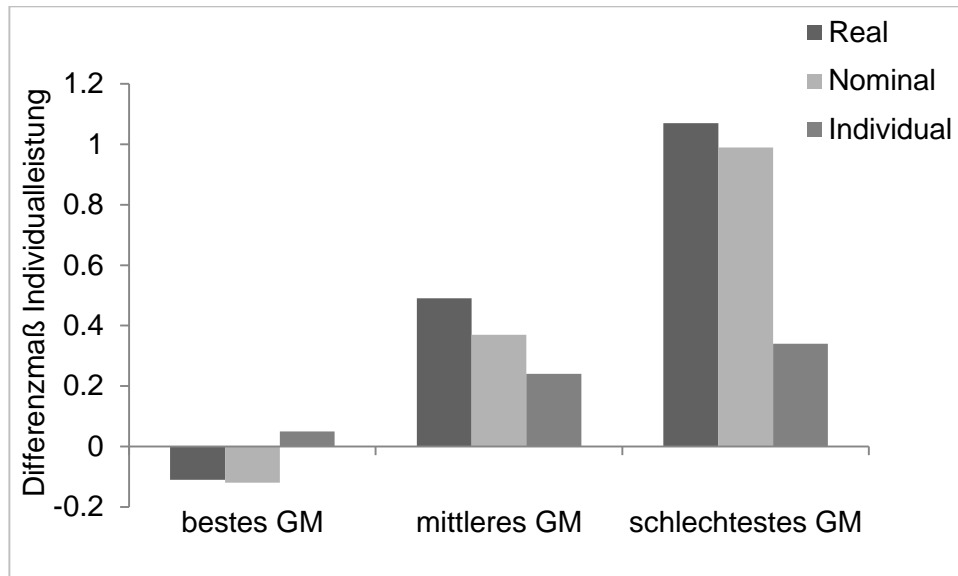


Abbildung 9. Ausmaß der individuellen Leistungssteigerung in Experiment 1 für das gemäß der Individualleistung in Phase 1 beste, mittlere und schlechteste Gruppenmitglied.<sup>32</sup>

Der Haupteffekt für den Faktor Gruppenmitglied war nicht signifikant,  $F(2, 202) = 1.72$ ,  $p = .182$ ,  $\eta^2_p = .02$ . Allerdings zeigte sich ein signifikanter Interaktionseffekt von Gruppenmitglied und Bedingung,  $F(2, 202) = 19.03$ ,  $p < .001$ ,  $\eta^2_p = .27$ .<sup>33</sup> Die in Abbildung 9 veranschaulichte Interaktion resultierte daraus, dass der zu erwartende Effekt, dass die schlechtesten Gruppenmitglieder besonders viel lernen, die anfangs mittelmäßigen moderat dazulernen und die anfangs besten nichts dazulernen, in der Realbedingung und der Nominalbedingung stärker auftrat als in der Individualbedingung, wo sich alle Gruppenmitglieder im Mittel nur geringfügig steigerten. Die Variabilität der individuellen Leistungszuwächse in der Individualbedingung war dabei, wie oben dargestellt, annähernd vergleichbar zu den anderen Bedingungen. Für eine genauere Analyse des individuellen Fertigkeitsszuwachses in Abhängigkeit vom Status der Gruppenmitglieder wurde jede

<sup>32</sup> Die individuelle Leistungssteigerung wurde über ein individuelles Differenzmaß aus dem MAPE der Phase 1 und dem MAPE der Phase 2 bestimmt.

<sup>33</sup> Die letzte mögliche Interaktion der ANCOVA von Gruppenmitglied und Startkompetenz war nicht signifikant,  $F(2, 202) = 1.93$ ,  $p = .15$ ,  $\eta^2_p = .02$ .

Bedingung individuell betrachtet. Die Ergebnisse der entsprechenden ANCOVAS finden sich in Anhang D.

Zusammenfassend kann aus den Ergebnissen abgeleitet werden, dass die Systemsteuerung in Realgruppen positive Auswirkungen auf die individuelle Steuerungskompetenz der Mitglieder hatte. Der Leistungszuwachs der Realgruppenmitglieder war den Nominalgruppenmitgliedern deskriptiv überlegen, wobei der Unterschied knapp nicht signifikant war. Dies Ergebnis könnte darauf hindeuten, dass durch den Kontext der Gruppenmittelwertbildung bzw. durch die Interaktion in der Gruppe ein Lerntransfer von der Gruppe auf das Individuum stattfand. Die individuelle Leistungssteigerung kann jedoch aus methodischen Gründen nicht eindeutig als G-I Transfer interpretiert werden: In der Nominal- und in der Realbedingung gaben die Gruppenmitglieder ihre individuellen Steuerungsurteile in jedem Trial auf Basis eines Systemzustandes ab, der aus den vorangehenden gruppenbasierten Urteilen (auf Basis der automatisch generierten Gruppenmittelwerte bzw. der Konsensusurteile der Realgruppen) resultierte. Aufgrund des assembly effect bonus (vgl. Collins & Guetzkow, 1964) resultierten durch die gruppenbezogene Steuerung günstigere Systemzustände, von denen die Real- und Nominalgruppenmitglieder auch für ihre individuelle Steuerung profitieren konnten. Die Probanden in der Individualbedingung hatten hingegen in ihren individuellen Steuerungsverläufen keinen situationsbegünstigenden statistischen Aggregationsvorteil. Der individuelle Lerneffekt in der Realbedingung beinhaltet somit sowohl einen statistischen Aggregationsvorteil als auch G-I Transfer. Das Ausmaß an G-I Transfer kann daher in Experiment 1 nicht bestimmt werden. Ein eindeutiger, methodisch sauberer Nachweis von G-I Transfer erfordert ein experimentelles Design, in dem die individuelle Steuerungsleistung *nach* der Gruppenphase mit einem individuellen Steuerungsverlauf erfasst wird und mit der individuellen Steuerungsleistung vor der Gruppenphase in Beziehung gesetzt wird. Experiment 3 beinhaltet dieses erforderliche Design zur methodisch sauberen Untersuchung des Auftretens von G-I Transfer, der über die Gruppenphase hinaus stabil ist.

#### **4.4.4 Wissenserwerb bei der Systemsteuerung**

Das verbalisierbare Wissen der Probanden über die Eigenschaften und Funktionsweise des Kühlsystems wurde jeweils nach der Explorationsphase, sowie nach der ersten und nach der zweiten Steuerungsphase erfasst. Wie unter Punkt 4.3.3 detailliert beschrieben, wurden die expliziten Wissensinhalte von zwei unabhängigen Kodierern kategorisiert und es wurde die Gesamtsumme an Annahmen, die Summe richtiger

Annahmen sowie der Prozentsatz richtiger Annahmen berechnet. Die Interraterreliabilität für die Summenscores bezüglich der Kodierung *richtig* vs. *falsch* lag bei  $Kappa = .81$  ( $SE = .01$ ),  $p < .001$ , 95% CI [.80, .83]. Nach der Kategorisierung von Landis und Koch (1977) ist dieser Übereinstimmungswert als *almost perfect agreement* zu bewerten. Die Interraterreliabilität für die inhaltliche Kodierung (die ebenfalls anschließend den Kategorien richtig vs. falsch zugeordnet und pro Proband summiert wurde) betrug  $Kappa = .67$  ( $SE = .01$ ),  $p < .001$ , CI [.65, .69]. Hierbei handelt es sich nach Landis und Koch (1977) um eine substantielle Übereinstimmung (*substantial agreement*).<sup>34</sup> Die Höhe der Interraterreliabilitäten rechtfertigte die anschließend durchgeführte Mittelung über die Werte der zwei Kodierer.

Es wurde überprüft, ob Realgruppen mehr verbalisierbares Wissen über die Systemeigenschaften erwerben als Nominalgruppen mit gemitteltem Systemverlauf und als Nominalgruppen mit individuellem Systemverlauf. Entsprechend der Teilhypothesen 4a bis 4c wurde dabei die Gesamtzahl an Annahmen, die Anzahl richtiger Annahmen und der Prozentsatz richtiger Annahmen betrachtet. Obwohl durch die Gesamtzahl an Annahmen und durch den Prozentsatz richtiger Annahmen die Quantitäts- und Qualitätsinformationen des Wissens abgedeckt sind, wird die Variable Anzahl richtiger Annahmen gesondert berichtet, um unmittelbar die durchschnittliche Häufigkeit richtiger Wissensinhalte der Probanden und somit das Ausmaß korrekten Systemwissens ablesen und vergleichen zu können.

Für die statistische Analyse wurde pro Proband jeweils ein Differenzmaß aus der ersten und zweiten Steuerungsphase für jede der drei abhängigen Variablen berechnet. Anschließend wurde der Mittelwert auf Gruppenebene bestimmt. Das Differenzmaß berücksichtigt Unterschiede im Ausgangswissen nach der ersten individuellen Steuerungsphase und ermöglicht Aussagen darüber, inwieweit sich die Quantität bzw. die Korrektheit des Wissens durch die Gruppeninteraktion im Vergleich zur Nominal- und Individualbedingung verändert hat. Eine ANOVA mit dem Zwischensubjektfaktor Bedingung (Realbedingung vs. Nominalbedingung vs. Individualbedingung) und der abhängigen Variablen Differenzmaß der Gesamtzahl an Annahmen ergab keinen signifikanten Bedingungsunterschied,  $F(2,102) = .20$ ,  $p = .822$ ,  $\eta^2_p = .00$ . In allen drei Bedingungen reduzierte sich die Gesamtzahl genannter Annahmen nach der zweiten Steuerungsphase im Vergleich zur Anzahl der Nennungen nach der ersten Steuerungsphase. Die Hypothese

---

<sup>34</sup> Landis und Koch (1977) interpretieren den Kappa-Koeffizienten in folgenden Kategorien:  $< 0$  *Poor agreement*, 0.0 – 0.20 *Slight agreement*, 0.21 – 0.40 *Fair agreement*, 0.41 – 0.60 *Moderate agreement*, 0.61 – 0.80 *Substantial agreement* und 0.81 – 1.00 *Almost perfect agreement*.

4a, die besagt, dass Personen in Realgruppen mehr Annahmen über die Systemeigenschaften treffen als Nominalgruppen mit gemitteltem oder individuellem Systemverlauf, wird daher verworfen. In Tabelle 4 sind die entsprechenden Mittelwerte und Standardabweichungen für die Bedingungen aufgeführt. Die Gesamtzahl an geäußerten Annahmen in der Realbedingung war nach der Gruppeninteraktion deskriptiv höher als in der Nominalbedingung, aber niedriger als in der Individualbedingung. Bei der Anzahl richtiger Annahmen ( $M = .94$ ,  $SD = .48$ ) sowie dem Prozentsatz richtiger Annahmen ( $M = 68.55$ ,  $SD = 30.44$ ) waren die Realgruppen den Vergleichsbedingungen hingegen deskriptiv überlegen

Die ANOVA mit dem Zwischensubjektfaktor Bedingung und der abhängigen Variablen Differenzmaß der Anzahl richtiger Annahmen war hingegen signifikant,  $F(2,102) = 8.84$ ,  $p < .001$ ,  $\eta^2_p = .15$ . Wie in Tabelle 4 ersichtlich, erhöhte sich im Mittel die Summe richtiger Annahmen in der Realbedingung, wohingegen sie sich in der Nominalbedingung reduzierte und in der Individualbedingung auf vergleichbarem Niveau verblieb<sup>35</sup>. Bonferroni-korrigierte Einzelvergleiche zeigten einen signifikanten Unterschied zwischen der Realbedingung und der Individualbedingung,  $p < .001$ , und einen knapp nicht signifikanten Unterschied zwischen der Real- und Nominalbedingung,  $p = .070$ . Nominal- und Individualbedingung unterschieden sich nicht signifikant,  $p = .186$ . Aufgrund des nicht signifikanten Nominalgruppenvergleichs bei strenger zweiseitiger Testung wird die Hypothese 4b nicht angenommen. Zwar fällt die *Anzahl korrekter Annahmen* über Systemeigenschaften bei Realgruppenmitgliedern hypothesenkonform höher aus als bei Mitgliedern von Nominalgruppen mit individuellem Systemverlauf, sie ist den Nominalgruppen mit gemitteltem Systemverlauf aber nur deskriptiv überlegen.

Schließlich wurde die Hypothese 4c überprüft, die besagt, dass der Prozentsatz korrekter Annahmen über Systemeigenschaften bei Realgruppenmitgliedern höher ausfällt als bei Mitgliedern von Nominalgruppen mit gemitteltem oder individuellem Systemverlauf. Hierzu wurde eine ANOVA mit dem Zwischensubjektfaktor Bedingung und der abhängigen Variablen Differenzmaß des Prozentsatzes richtiger Annahmen durchgeführt, die einen signifikanten Bedingungsunterschied ergab,  $F(2,71) = 3.98$ ,  $p = .023$ ,  $\eta^2_p = .101$ . Bonferroni-korrigierte paarweise Vergleiche bestätigten eine signifikante Überlegenheit der

---

<sup>35</sup> Separate Ein-Stichproben- $t$ -Tests gegen 0 ergaben für das Differenzmaß der Anzahl richtiger Annahmen in der Realbedingung einen signifikanten Anstieg richtiger Annahmen,  $t(34) = 2.12$  und  $p = .041$ , für die Nominalbedingung keine signifikante Veränderung,  $t(34) = 0.88$  und  $p = .383$  und für die Individualbedingung mit  $t(34) = 3.65$  und  $p < .001$  eine signifikante Abnahme der Anzahl richtiger Annahmen.

Realbedingung gegenüber der Individualbedingung,  $p = .001$ . Der Prozentsatz richtiger Annahmen unterschied sich hingegen nicht signifikant zwischen der Real- und der Nominalbedingung,  $p = .508$ , sowie zwischen der Nominal- und der Individualbedingung,  $p = .686$ . Somit kann die Hypothese 4c nur in Bezug auf den größeren Prozentsatz richtigen Systemwissens von Realgruppenmitgliedern im Vergleich zu individuell steuernden Probanden gestützt werden. Die Teilvorhersage zur Unterlegenheit von computer- vernetzten Nominalgruppenmitgliedern gegenüber Realgruppenmitgliedern hat sich in dieser Analyse nicht bestätigt.

Tabelle 4: *Mittelwerte und Standardabweichungen der drei abhängigen Variablen Gesamtzahl der Annahmen, Anzahl richtiger Annahmen und Prozentsatz richtiger Annahmen für die absoluten Werte in Steuerungsphase 2 sowie für das Differenzmaß aus Phase 1 und Phase 2*

Variable	<i>n</i>	Werte Phase 2 <i>M (SD)</i>	<i>n</i>	Differenzmaß <i>M (SD)</i>
<b>Gesamtzahl Annahmen</b>				
Realbedingung	35	1.39 (.53)	35	-.27 (1.12)
Nominalbedingung	35	.88 (.37)	35	-.38 (.99)
Individualbedingung	35	1.61 (.54)	35	-.40 (.78)
<b>Anzahl richtig</b>				
Realbedingung	35	.94 (.48)	35	.33(.92)
Nominalbedingung	35	.40 (.34)	35	-.10 (.67)
Individualbedingung	35	.49 (.39)	35	-.45 (.73)
<b>Prozentsatz richtig</b>				
Realbedingung	34	68.55 (30.44)	26	28.05 (53.33)
Nominalbedingung	34	35.38 (.24.39)	20	7.08 (51.53)
Individualbedingung	35	33.60 (28.31)	28	-11.34 (50.61)

*Anmerkung: Die unvollständigen Zellbesetzungen bei der Variablen Prozentsatz richtig resultierten daraus, dass in den fehlenden Fällen keine Annahmen nach der Steuerungsphase 2 getroffen wurden.*

In einer Mediationsanalyse wurde überprüft, ob die (gegenüber der Individualbedingung signifikant und gegenüber der Nominalbedingung deskriptiv) bessere Leistung der Realgruppen durch G-I Transfer in Form eines zu größeren Anteilen korrekten Systemwissens erklärt werden konnte.<sup>36</sup> Hierzu wurde als Maß für den wissensbezogenen G-I Transfer die Differenz des Prozentsatzes korrekter Systemannahmen aus Phase 2 und Phase 1 berechnet. Positive Werte des Differenzmaßes bedeuten eine Steigerung der Wissensqualität in Phase 2. Damit nach Baron und Kenny (1986) eine Mediation nachgewiesen ist, muss gezeigt werden, dass (1) die Gruppenbedingung einen signifikanten Effekt auf die Gruppenleistung hat, (2) die Gruppenbedingung einen signifikanten Effekt auf den Mediator des wissensbezogenen G-I Transfers hat, und (3) dass der Effekt der Bedingung auf die Gruppenleistung signifikant reduziert wird, wenn für wissensbezogenen G-I Transfer kontrolliert wird. Der wissensbezogene G-I Transfer sollte hingegen seinen signifikanten Einfluss auf die Leistung beibehalten, wenn gleichzeitig die Bedingung in die Regressionsgleichung eingeht. Für die Berechnungen wurden die Bedingungen dummy-kodiert in die zwei Kategorien Realgruppenbedingung vs. keine Realgruppenbedingung. Letztere Kategorie beinhaltete die Nominalgruppen mit gemitteltem und mit individuellem Systemverlauf. Die unabhängige Variable der Gruppenzugehörigkeit hatte sowohl einen signifikanten Effekt auf die Gruppenleistung,  $\beta = -.46$ ,  $t(54) = -4.426$ ,  $p < .001$ , als auch auf den wissensbezogenen G-I Transfer,  $\beta = .36$ ,  $t(54) = 2.82$ ,  $p = .007$ . Bei gleichzeitigem Einfügen der Gruppenzugehörigkeit und des G-I Transfers war G-I Transfer weiterhin ein signifikanter Prädiktor für die Gruppenleistung,  $\beta = -.34$ ,  $t(54) = -2.55$ ,  $p = .007$ . Der Zusammenhang zwischen der unabhängigen Variablen und der Gruppenleistung war hingegen nicht mehr statistisch signifikant,  $\beta = -.25$ ,  $t(54) = -1.87$ ,  $p = .06$ . Die Reduktion des Regressionsgewichtes für die Variable Gruppenzugehörigkeit war statistisch signifikant, wie ein durchgeführter Sobel Test zeigte (vgl. Sobel, 1982),  $z = -1.86$ .  $p = .003$  bei einseitiger bzw.  $p = .063$  bei zweiseitiger Testung. In diesen explorativen Berechnungen finden sich also Hinweise darauf, dass der positive Effekt des Realgruppenkontextes auf die Gruppenleistung durch den wissensbezogenen G-I Transfer mediiert worden sein könnte. Für den zweiten möglichen Mediator des wissensbezogenen G-I Transfers in Form der Differenz der Anzahl richtiger Annahmen aus

---

<sup>36</sup> Eine Mediationsanalyse für den Bedingungsvergleich zwischen der Realbedingung und der Nominalbedingung ist statistisch unzulässig, da sich die abhängige Variable Steuerungsleistung zwischen diesen Bedingungen nicht signifikant unterscheidet. Die Realbedingung wurde mit den beiden anderen Bedingungen in einer dichtotomen UV kontrastiert. Somit ist diese Mediationsanalyse lediglich als explorative Auswertung zu interpretieren

Phase 2 und Phase 1 zeigten sich hingegen keine signifikanten Effekte, die auf eine Mediation hindeuten.<sup>37</sup>

Zuletzt wurde explorativ betrachtet, ob es einen Zusammenhang zwischen der erbrachten Leistung in der zweiten Steuerungsphase und den Indikatoren für Systemwissen gab. Die Quantität des Wissens hing nicht mit der Steuerungsleistung zusammen. Die Korrelation zwischen dem MAPE und der Gesamtsamtsumme der Annahmen in Phase 2 betrug  $r = -.06$ ,  $p = .284$ . Zwischen der Anzahl richtiger Annahmen und dem MAPE gab es hingegen einen signifikanten Zusammenhang,  $r = -.39$ ,  $p < .001$ . Auch der Prozentsatz richtiger Annahmen wies mit  $r = -.37$ ,  $p = .001$ , einen signifikanten Zusammenhang zur Steuerungsleistung auf. Da beim Leistungsmaß MAPE niedrige Werte eine bessere Leistung implizieren, ging eine höhere Wissensqualität somit mit höherer Steuerungsleistung einher. In Einklang mit einigen vorausgehenden Forschungsbefunden (Kröner, 2001; Putz-Osterloh, Bott, & Houben, 1988) lag ein signifikanter Zusammenhang zwischen der Anzahl bzw. dem Prozentsatz richtiger Systemannahmen und der vorausgehenden Steuerungsleistung vor. Darüber hinaus entsprach die Abstufung der Wissensqualität mit einem abfallenden Prozentsatz richtiger Wissensinhalte von der Realbedingung über die Nominalbedingung zur Individualbedingung der deskriptiven Leistungsrangreihe zwischen den Bedingungen. In Bezug auf richtige Wissensinhalte zu den Variablenzusammenhängen und der Funktionsweise des Kühlsystems trat also G-I Transfer in der Realgruppenbedingung auf.

#### **4.4.5 Explorative Analysen: Erkennen von steuerungsbezogener Expertise und subjektive Gewichtung der Individualurteile im Realgruppenkontext**

Die Analysen zu Lerneffekten auf Gruppenebene haben gezeigt, dass es Realgruppen nicht gelang, Koordinationsgewinne bei der Steuerung des dynamischen

---

<sup>37</sup> Die unabhängige Variable der Gruppenzugehörigkeit hatte sowohl einen signifikanten Effekt auf die Gruppenleistung,  $\beta = -.28$ ,  $t(104) = -2.95$ ,  $p = .004$ , als auch auf den wissensbezogenen G-I Transfer,  $\beta = .34$ ,  $t(104) = 3.71$ ,  $p < .001$ . Bei gleichzeitigem Einfügen der Gruppenzugehörigkeit und des G-I Transfers, war G-I Transfer kein signifikanter Prädiktor für die Gruppenleistung mehr,  $\beta = -.15$ ,  $t(104) = -1.47$ ,  $p = .146$ . Der Zusammenhang zwischen der unabhängigen Variablen und der Gruppenleistung war hingegen weiterhin statistisch signifikant,  $\beta = -.23$ ,  $t(104) = -2.28$ ,  $p = .025$ . Somit waren die Voraussetzungen für eine Mediation für das Differenzmaß der Anzahl richtiger Annahmen nicht erfüllt. Die Reduktion des Regressionsgewichtes für die Variable Gruppenzugehörigkeit war statistisch entsprechend nicht statistisch signifikant,  $z = 1.35$ .  $p = .088$  bei einseitiger bzw.  $p = .176$  bei zweiseitiger Testung.



Kühlsystemszenarios zu realisieren. Die erzielte Gruppenleistung unterschied sich nicht signifikant vom Gruppenpotenzial, welches als die Steuerungsleistung definiert wurde, die aus der Mittelung der individuellen Steuerungsurteile pro Trial resultiert hätte. Ein Vergleich der Gruppenleistung mit dem True Best Member Model (der Individualleistung des besten Gruppenmitglieds in Phase 2) zeigte, dass die Gruppe signifikant schlechter abschnitt. Eine Übernahme der Individualurteile des besten Gruppenmitglieds als Gruppenurteile hätte folglich zu einer signifikanten Steigerung der Gruppenleistung geführt. Damit die Realgruppenurteile von einer 100%igen oder zumindest überdurchschnittlichen (über 33.33% liegenden) Gewichtung des besten Gruppenmitglieds profitieren könnten, ist das Erkennen von Unterschieden in der Steuerungsexpertise zwischen den Gruppenmitgliedern eine notwendige Voraussetzung. Im Folgenden soll im ersten Schritt geprüft werden, ob die Realgruppenmitglieder überzufällig häufig das Expertiselevel der individuellen Steuerung in Phase 2 (bzw. Phase 1) erkannten und das beste und schlechteste Mitglied ihrer Gruppe identifizieren konnten. Im zweiten Schritt wird untersucht, ob die Realgruppenmitglieder die Individualurteile nach eigener Einschätzung gemäß wahrgenommener Expertiseunterschiede gewichteten, um ihre Gruppenurteile festzulegen.

Die Probanden schätzten nach der zweiten Steuerungsphase unabhängig voneinander die eigene sowie die Kompetenz der zwei anderen Realgruppenmitglieder anhand eines relativen und eines absoluten Maßes ein. Die relative Kompetenzbeurteilung erfolgte, indem alle Gruppenmitglieder in eine Rangreihe gebracht wurden und die drei Aussagen „Das kompetenteste (bzw. das zweitkompetenteste bzw. das am wenigsten kompetente) Mitglied in meiner Gruppe war Mitglied Nummer...“ ergänzten. Die absolute Kompetenzeinschätzung jedes Gruppenmitglieds erfolgte auf einer Skala von 1 (*sehr geringe Kompetenz*) bis 7 (*sehr hohe Kompetenz*).

Die richtige Bestimmung des besten und schlechtesten Mitglieds wurde für die relative Kompetenzeinordnung anhand einer dichotomen Variablen bestimmt, wobei eine korrekte Zuordnung als bestes bzw. schlechtestes Mitglied mit dem Wert 1 und eine falsche Zuordnung mit 0 rekodiert wurde. Anschließend wurde der Gruppenmittelwert für die dichotome Variable berechnet. Bei Identifikation des besten (bzw. schlechtesten) Mitglieds durch alle drei Gruppenmitglieder resultierte ein Wert von 1, bei richtiger Identifikation durch zwei Gruppenmitglieder ein Wert von .67 und erkannte nur ein Gruppenmitglied das beste (bzw. schlechteste) Mitglied resultierte ein Wert von .33. Falls kein Gruppenmitglied die richtige Zuordnung traf, ergab sich ein Wert von 0. Der gruppenbasierte Identifikationswert wurde gegen den Erwartungswert .33 unter zufälliger Auswahl eines Mitglieds getestet. Die

Probanden erkannten in 44% der Fälle und damit nicht überzufällig häufig das nach der Individualleistung beste Gruppenmitglied (true best member) in der zweiten Steuerungsphase,  $t(30) = 1.52$ ,  $p = .140$ ,  $d = 0.26$ . Auch das individuell schlechteste Gruppenmitglied (true worst member) in der Steuerungsphase 2 wurde lediglich in 30% der Fälle und damit nicht überzufällig häufig identifiziert,  $t(30) = -.49$ ,  $p = .625$ ,  $d = 0.10$ . Der prozentuale Anteil richtiger Kompetenzeinschätzungen für das beste und schlechteste Mitglied ist in Tabelle 5 dargestellt.

Es ist denkbar, dass die Gruppenmitglieder ihre Beurteilung der relativen Steuerungskompetenz anhand des ersten Eindrucks zu Beginn der zweiten Steuerungsphase fassten, beispielsweise durch Bewertung des anfänglich geäußerten Systemverständnisses oder der Bewertung der individuellen Steuerungsstrategie und dem zugehörigen Steuerungserfolg in der ersten Phase. Im Folgenden wurde untersucht, ob es den Realgruppenmitgliedern gelang, überzufällig häufig gemäß der Leistung in der ersten individuellen Steuerungsphase das beste (initial best member) bzw. schlechteste Mitglied (initial worst member) zu bestimmen. Tatsächlich konnten die Gruppen wie in Tabelle 5 zu sehen in 61% ( $SD = 36.61$ ) der Fälle mit überzufälliger Wahrscheinlichkeit das initial best member identifizieren,  $t(30) = 4.30$ ,  $p < .001$ ,  $d = 0.75$ . Das schlechteste Mitglied der individuellen Steuerungsphase wurde mit 46% ( $SD = 38.18$ ) zwar deskriptiv, aber nicht überzufällig häufig richtig erkannt,  $t(30) = 1.93$ ,  $p = .063$ ,  $d = 0.33$ .

Tabelle 5: Mittelwerte und Standardabweichungen des Prozentsatzes richtig identifizierter relativer Mitgliederkompetenz in der Realbedingung

Bedingung	<i>n</i>	<i>M (SD)</i>
Initial best member	35	.61 (.37)
Initial worst member	35	.46 (.38)
True best member	35	.44 (.41)
True worst member	35	.30 (.33)

Die erzwungene klare Vergabe der Rangplätze bei der relativen Kompetenzeinschätzung ermöglichte keine Einschätzung darüber, wie groß die wahrgenommenen Unterschiede innerhalb der subjektiven Kompetenzrangreihe der Probanden waren. In einer weiteren Analyse wurde daher exploriert, wie groß die

quantitativen Unterschiede in der *absoluten* Kompetenzeinschätzung der Gruppenmitglieder auf der Skala von 1 (*sehr geringe Kompetenz*) bis 7 (*sehr hohe Kompetenz*) waren. Hierzu wurde das pro Gruppe gemittelte Kompetenzrating für das beste, mittlere und schlechteste Gruppenmitglied gemäß der Individualleistung in Phase 2 berechnet. In einer MANOVA mit dem Innersubjektfaktor Gruppenmitglied (true best member vs. true medium member vs. true worst member) zeigte sich kein Unterschied im mittleren Kompetenzrating der Realgruppen,  $F(2,64) = 1.01$ ,  $p = .370$ ,  $\eta^2_p = .03$ . In Einklang mit der nicht überzufällig korrekten Leistungsrangfestlegung des besten und schlechtesten Mitglieds fiel die mittlere Kompetenzeinschätzung des true best member mit 5.17 ( $SD = .93$ ) sehr ähnlich zur Einschätzung des true medium member ( $M = 4.94$ ,  $SD = .97$ ) sowie des true worst member ( $M = 4.85$ ,  $SD = 1.04$ ) aus. In 4 der 35 Realgruppen war das mittlere Kompetenzrating für alle drei Gruppenmitglieder sogar identisch. Die vergleichbare MANOVA mit dem Innersubjektfaktor Gruppenmitglied gemäß der Individualleistung in Phase 1 (initial best vs. initial medium vs. initial worst member) war hingegen signifikant,  $F(2,64) = 11.71$ ,  $p < .001$ ,  $\eta^2_p = .27$ . Paarweise post hoc Tests zeigten eine signifikant höhere Kompetenzeinschätzung des initial best member ( $M = 5.55$   $SD = .13$ ) als des initial worst ( $M = 4.68$   $SD = .16$ ) mit  $p < .001$ , sowie des initial medium member auf ( $M = 4.73$   $SD = .18$ ),  $p = .002$ . Das Kompetenzrating zwischen initial medium und worst member unterschied sich hingegen nicht,  $p = .799$ .

Nach der Betrachtung des relativen und absoluten Kompetenzratings der Realgruppenmitglieder wird nun in zwei Analysen exploriert, ob die subjektive Gewichtung<sup>38</sup> der Individualurteile in der Gruppenphase differenziert gemäß der subjektiven (und wie oben dargestellten häufig falschen) Expertiserangreihe bzw. gemäß der tatsächlichen Leistungsangreihe erfolgte. Eine erste MANOVA mit dem Innersubjektfaktor subjektive Gruppenmitgliedsexpertise (subjective best member vs. subjective medium member vs. subjective worst member) und der abhängigen Variable der pro Gruppe gemittelten subjektiven Gewichtung pro Mitglied zeigte einen signifikanten Gewichtungsunterschied zwischen den Gruppenmitgliedern,  $F(2, 56) = 9.47$ ,  $p < .001$ ,  $\eta^2_p = .25$ . Die

---

<sup>38</sup> Die Probanden schätzten nach der Realgruppeninteraktion ein, wie stark die individuellen Urteile pro Gruppenmitglied über alle Trials hinweg in das Gesamturteil der Systemsteuerwerte Ihrer Gruppe eingeflossen seien. Dabei machten Sie prozentuale Angaben, die sich für alle drei Einschätzungen zu 1 bzw. 100% aufaddierten. Als abhängige Variable in der MANOVA wurde die mittlere subjektive Gewichtung aller drei Probanden pro Gruppenmitglied berechnet. Das Gewichtungsmaß basierte also auf einer Selbst- und zwei Fremdeinschätzungen.

Realgruppenmitglieder gewichteten nach eigener Einschätzung die Individualurteile des subjective best member ( $M = 39.10$ ,  $SD = 9.15$ ) signifikant stärker ( $p = .042$ ) als die des subjective medium member ( $M = 31.36$ ,  $SD = 6.78$ ) und mit  $p = .002$  signifikant stärker als die des subjective worst member ( $M = 28.16$ ,  $SD = 7.77$ ).<sup>39</sup> Die wahrgenommene Gewichtung zwischen dem subjective medium und worst member war nicht signifikant,  $p = .155$ . Damit wichen die Probanden entsprechend ihrer subjektiven Kompetenzeinschätzung bewusst von einer Mittelungsstrategie der Individualurteile zur Gruppenurteilsbildung ab und gaben nach eigenen Angaben den Beiträgen des vermeintlich kompetentesten Gruppenmitglieds mehr Gewicht.

Eine zweite MANOVA mit dem Innersubjektfaktor wahre Gruppenmitgliedsexpertise (true best member vs. true medium member vs. true worst member) und der pro Gruppe gemittelten subjektiven Gewichtung pro Mitglied als abhängiger Variable war hingegen nicht signifikant,  $F(1.69, 53.94) = 1.23$ ,  $p = .295$ ,  $\eta^2_p = .04$ . Die Realgruppenmitglieder gewichteten subjektiv die Individualurteile des true best member ( $M = 33.08$ ,  $SD = 8.33$ ), des true medium member ( $M = 31.36$ ,  $SD = 6.78$ ) und des true worst member ( $M = 35.31$ ,  $SD = 9.84$ ) in etwa gleich stark. Damit entspricht die subjektive Gewichtung der Individualurteile der unter Punkt 4.4.2 festgestellten empirischen Mittelwertbildung zur Gruppenurteilsfindung (bzw. der Passung des Average Models auf die Daten). Da die Gruppen nicht überzufällig gut die korrekte Kompetenzrangreihe der Gruppenmitglieder erkennen konnten (siehe oben), intendierten sie folglich nicht das beste Gruppenmitglied stärker zu gewichten, obwohl generell die Bereitschaft vorhanden war, dem besten Mitglied mehr Gewicht zu geben.

Bei der dynamischen Kühlsystemsteuerungsaufgabe im Atomkraftwerkszenario war das initial best member nur in 45.71% der Fälle ( $SD = 50.54$ ) auch in Bezug auf die Individualurteile in der Gruppenphase das beste Mitglied. Dieser Prozentsatz war nicht von

---

<sup>39</sup> Die subjektiv angegebenen Gewichtungen sind lediglich ein Indikator für die angestrebte Gewichtung. Auch bleibt unklar, ob der pauschale Gewichtungswert der Probanden einen subjektiven Mittelwert der Mitgliedsgewichtung über alle 60 Trials der Phase 2 darstellt oder die Probanden tatsächlich einen konstanten Gewichtungsfaktor pro Gruppenmitglied über alle Trials wahrnahmen. In zukünftigen Studien wäre es interessant, die wahrgenommene Variabilität der subjektiven Gewichtung zu erfragen.

Bei statischen Urteilsaufgaben wäre es in diesem Fall sinnvoll als Indikator für die Korrektheit der subjektiv vorgenommenen Gewichtung die Passung eines Gewichtungsmodells nach der von der Gruppe angegebenen Gewichtung auf die Daten zu überprüfen. Bei der dynamischen Steuerungsaufgabe ist dies jedoch wenig sinnvoll, da post hoc neuartige Systemverläufe auf Basis der gemittelten subjektiven Gewichtungswerte berechnet werden müssten und die Abweichungen der daraus resultierenden MAPE-Werte von den Gruppenwerten nicht eindeutig interpretierbar wären.

der Zufallswahrscheinlichkeit von 33% verschieden,  $t(34) = 1.49$ ,  $p = .146$ , 95% CI [-.47, .30]. Das initial worst member war hingegen in 54.29% der Fälle ( $SD = 50.54$ ) und damit überzufällig häufig auch während der zweiten Steuerungsphase das schlechteste Gruppenmitglied,  $t(34) = 2.49$ ,  $p = .018$ , 95% CI [.04, .39].

Eine mögliche Erklärung für die mangelhafte Kompetenzerkennung in den ad hoc zusammengestellten Realgruppen ist die Kompetenzeinschätzung nach Extraversion der Gruppenmitglieder. Vorangehende Studien konnten zeigen, dass der Redeanteil und somit auch der potenzielle soziale Einfluss von extravertierten Gruppenmitgliedern höher war (Littlepage, Schmidt, Whisler, & Frost, 1995) und dass Gruppen bei Aufgaben mit eingeschränkter Demonstrierbarkeit der richtigen Lösung extravertiertere Mitglieder stärker gewichteten (Bonner, 2000; Bonner & Sillito, 2011; Bonner, Sillito, & Baumann, 2007). Auch das dynamische Kühlsystem weist eine intransparente Struktur auf und geeignete Lösungsstrategien sind wegen der zeitverzögerten Systemreaktion und der Oszillation der Temperaturwerte nur bedingt demonstrierbar. Im Gegensatz zu den bisherigen Befunden bei visuellen Entscheidungsaufgaben (Bonner, 2000) und bei quantitativen Schätzaufgaben (Bonner, Sillito, & Baumann, 2007; Bonner & Sillito, 2011) erfolgte die Kompetenzbeurteilung der Gruppenmitglieder in Experiment 1 jedoch nicht nach Extraversionsausprägung. Wie unter Abschnitt 4.4.2 gezeigt wurde, wies das normative Extraversionmodell eine schlechte Passung auf die Realgruppendaten auf. Auch die subjektiv berichtete und die empirische Gewichtung der Beiträge orientierte sich nicht an der Extraversionsausprägung (vgl. Anhang E).

Zusammenfassend können drei Ergebnisse zum Erkennen von steuerungsbezogener Expertise und zu der subjektiven Gewichtung der Individualurteile in interagierenden Gruppen im Kühlsystemszenario festgehalten werden: Erstens erkannten die Probanden weder überzufällig häufig das beste noch das schlechteste Gruppenmitglied gemäß der individuellen Steuerungsurteile während der Gruppenphase, wenn sie sich und die anderen Gruppenmitglieder in eine Kompetenzrangreihe bringen sollten und die quantitative Kompetenzeinschätzung zwischen dem tatsächlich besten, mittleren und schlechtesten Gruppenmitglied in der Gruppenphase unterschied sich nicht signifikant. Zweitens betrachteten die Probanden überzufällig häufig das Gruppenmitglied mit der besten individuellen Leistung in der ersten Steuerungsphase als bestes Gruppenmitglied. Mit Bezug auf das erste Ergebnis kann man daraus schlussfolgern, dass es den Probanden nicht gelang, Änderungen in der Kompetenzrangreihe zwischen der ersten Individual- und der zweiten Gruppensteuerungsphase zu erkennen. Drittens gaben die Realgruppenmitglieder nach der Gruppenphase an, die individuellen Beiträge des subjektiv

besten Gruppenmitgliedes stärker für die Gruppenurteile gewichtet zu haben als die Beiträge der beiden anderen Gruppenmitglieder. Aufgrund der nicht überzufällig korrekten Erkennung des besten Gruppenmitglieds konnten die Gruppen allerdings nicht von ihrer differenzierten Gewichtungsabsicht profitieren.<sup>40</sup>

Darüber hinaus kann aus den Ergebnissen geschlossen werden, dass Leistungsfeedback zur Leistungsrangreihe gemäß der individuellen Steuerungsleistung vor der Gruppeninteraktion entgegen der in der Literatur häufig angenommenen positiven Effekte auf die Leistung (vgl. Henry, Strickland, Yorges, & Ladd, 1996; Moreland & Myaskovsky, 2000; Schultze et al., 2012) nur bedingt geeignet wäre, um das Erkennen von (nicht vorhandener) Expertise zu fördern und zu einer Gruppenleistungssteigerung beizutragen. Zwar könnten Gruppen bei der Kühlsystemsteuerung durch Leistungsfeedback überzufällig gut ihr während der Interaktion schlechtestes Mitglied erkennen, die Identifikation des besten Mitglieds, auf dessen stärkere Gewichtung es zur Leistungssteigerung ankommt, würde aber nicht überzufällig häufig erkannt. Falls also Leistungsfeedback zum Einsatz kommen sollte, wäre ergänzend der Hinweis an die Probanden sinnvoll, dass sich (insbesondere) der Rangplatz des besten Mitglieds bei der Abgabe der Individualurteile während der Gruppensteuerung ändern kann, sodass eine kritische Beurteilung nach Akkuratheit in den einzelnen Trials sinnvoll ist. Ob solch ein Leistungsfeedback mit Zusatzinformation einen leistungsförderlichen Effekt hat, der über die Instruktion jeweils das akkurateste Individualurteil pro Trial zu identifizieren hinausgeht, müsste empirisch überprüft werden.<sup>41</sup> Falls in anderen dynamischen Steuerungsszenarien die Leistungsrangreihe *aller* Gruppenmitglieder über die Gruppenphase stabil bliebe, sollte die Wahrscheinlichkeit für das Auftreten von Koordinationsgewinnen zunehmen und Leistungsfeedback würde als leistungsförderliche Intervention sinnvoll sein.

---

<sup>40</sup> Alternativ könnten die Gruppenmitglieder auch bewusst eine Mittelungsstrategie (unabhängig von den subjektiv wahrgenommenen Expertiseunterschieden innerhalb der Gruppe) während der Steuerung angewandt haben und im nachfolgenden Fragebogen bei der Angabe der Gewichtungsabsicht falsche Angaben gemacht haben um zu ihren subjektiven Expertiseeinschätzungen konsistentes Urteilsverhalten zu demonstrieren. Wenn man davon ausgeht, dass die Probanden nicht systematisch falsche Antworten gegeben haben, ist jedoch die erste im Text erläuterte Interpretation wahrscheinlicher.

<sup>41</sup> Die Realgruppendaten in Experiment 1 zeigten keinen signifikanten Leistungsunterschied unter Anwendung des Average Models im Vergleich zum True Best Member Models,  $t(34) = 1.83$ ,  $p = .075$ ,  $d = 0.04$ .

#### 4.4.6 Zusammenfassung und Ausblick auf Experiment 2

In Experiment 1 sollte erstmalig methodisch sauber untersucht werden, ob die Systemsteuerung durch Gruppen zu interaktionsbedingten Prozessgewinnen führt, die über den statistischen Aggregationsvorteil von Einzelurteilen hinausgehen. Zu diesem Zweck wurde eine neue Vergleichsbedingung eingeführt, die neben der in der Literatur etablierten Nominalgruppenbedingung mit individuellem Systemverlauf untersucht wurde. Hierbei waren räumlich getrennte Probanden über einen Computer zu Nominalgruppen mit gemitteltem Systemverlauf vernetzt. Pro Steuerungsdurchgang berechnete ein Computer automatisch den Mittelwert aus den individuellen Steuerungsurteilen der Probanden. Durch die statistische Aggregation wurde die Fehlerstreuung reduziert und es entstand ein statistischer Vorteil gegenüber der aus individuellen Steuerungsverläufen berechneten Nominalgruppenleistung in der Individualbedingung. Die Probanden steuerten jeweils zunächst in einer Individualphase und dann in einer zweiten Phase unter einer der drei Bedingungen einen einfachen, dynamischen Regelkreis (vgl. Reichert & Dörner, 1988). Dieser war in der vorliegenden Arbeit semantisch in ein Atomkraftwerkszenario mit einem zu optimierenden Kühlsystem eingekleidet. Als Leistungsmaß wurde der MAPE berechnet, der die mittlere prozentuale Abweichung der durch Steuerung erzielten Kühlsystemtemperatur vom anzustrebenden optimalen Zustand angab.<sup>42</sup>

Experiment 1 zeigte, dass die Gruppenleistung bei face-to-face Interaktion in der Realbedingung den Nominalgruppen mit individuellem Systemverlauf signifikant überlegen war. Die Steuerungsleistung der Realgruppen ging jedoch nur deskriptiv und nicht signifikant über den Leistungsgewinn durch statistische Urteilsaggregation in der Nominalbedingung mit gemitteltem Systemverlauf hinaus. Gemessen an diesem neu eingeführten, strengen Vergleichsmaßstab führte die Gruppeninteraktion per se also nicht zu signifikanten Prozessgewinnen. Zusammenfassend ist zum Experiment 1 festzuhalten, dass die Realgruppen zwar die durchschnittliche individuelle Steuerungsleistung übertrafen, aber durch gemeinsame Systemsteuerung keine signifikant bessere Leistung erzielten als durch den statistischen Aggregationsvorteil bei Mittelung der individuellen Beiträge in der Nominalbedingung resultierte. Den Realgruppenmitgliedern gelang es nicht, Koordinationsgewinne zu erzielen, indem sie ihr durchschnittlich während der

---

<sup>42</sup> Als Benchmarks zur Beurteilung der absoluten Steuerungsleistung pro Bedingung wurden die MAPE-Werte herangezogen, die unter Anwendung einer langfristigen Optimalstrategie mit der Stellradeinstellung 23 bzw. unter dem Nichteingreifen und Beibehaltung des Stellradausgangswertes resultierten. Die durchschnittliche Steuerungsleistung in allen Bedingungen lag innerhalb des Leistungsbereiches dieser Grenzen einer besonders guten bzw. schlechten Steuerungsleistung.

Gruppenphase bestes Mitglied oder das pro Trial akkuratete Urteil stärker gewichteten. Wie der Vergleich der Realgruppenleistung mit der hypothetischen Leistung unter dem True Best Member bzw. dem Accuracy Model zeigte, hätten beide Vorgehensweisen die Gruppenleistung signifikant steigern können. Die Realgruppenmitglieder intendierten zwar die Individualurteile des kompetentesten Gruppenmitglieds stärker zu gewichten, sie konnten den Kompetenzstatus allerdings nicht überzufällig gut feststellen. Somit resultierte eine (wahrscheinlich unbeabsichtigte) Mittelungsstrategie zur Bestimmung der Gruppenurteile.

Es ist unklar, inwieweit die Ergebnisse des Experimentes 1 auf andere Systemszenarien generalisierbar sind. Das dynamische Kühlsystem im Atomkraftwerkszenario wies mit nur einer Eingabevariablen zur Systemsteuerung (der Stellradeinstellung) und einer beobachtbaren Ausgabevariablen zur Systemreaktion (der resultierenden Temperatur) eine geringe Komplexität auf. Dies schränkte den Informationsgehalt der Gruppendiskussion über mögliche Variablenzusammenhänge, Nebeneffekte und Rückkopplungsmechanismen im System stark ein. Entsprechend gab es wenig argumentative Anhaltspunkte um während der Gruppenphase die Vorschläge des True Best Member bzw. das pro Trial akkuratete Individualurteil zu identifizieren. Es ist denkbar, dass es Gruppen besser gelingt, in einem System mit mehr beobachtbaren Variablen, das beste Mitglied bzw. die akkurateten Individualurteile zu erkennen, da eventuell bessere Möglichkeiten bestehen, Systemwissen anhand beobachtbarer Systemreaktionen zu prüfen. Solch eine erhöhte Demonstrierbarkeit sollte das Auftreten von Koordinationsgewinnen durch G-IG Transfer bei Gewichtung nach Akkuratheit und/oder G-G-Transfer durch Gewichtung des True Best Member ermöglichen. Die so erzielte Leistungssteigerung der Realgruppensteuerung würde dann mit hoher Wahrscheinlichkeit zu Prozessgewinnen führen, die nicht nur die Leistung von Nominalgruppen mit individuellem Systemverlauf sondern auch die Nominalgruppenleistung bei gemitteltem Systemverlauf übertreffen. Im folgenden Experiment soll daher untersucht werden, ob die Befunde aus Experiment 1 zum relativen Gruppenleistungsniveau und ausbleibenden Koordinationsgewinnen in einem komplexeren dynamischen System mit einer größeren Anzahl vernetzter, beobachtbarer Variablen repliziert werden können.



## 5. Experiment 2

### 5.1 Zielsetzung und Hypothesen

Die Zielsetzung des Experimentes 2 ist die Überprüfung der Replizierbarkeit der Befunde aus dem ersten Experiment mit einer komplexeren dynamischen Systemsimulation. Hierbei wurde ein System gewählt, das eine größere Anzahl beobachtbarer Systemvariablen aufweist. Dies ermöglicht den Erwerb differenzierteren Systemwissens durch beobachtbare Veränderungen multipler Variablenbeziehungen, sowie eine größere inhaltliche Diskussionsbasis zur Funktionsweise des Systems bei der Gruppensteuerung in der Realbedingung. Mehr beobachtbare Systemparameter sollten während der Gruppensteuerung die Demonstrierbarkeit richtigen Systemwissens, guter Steuerungsstrategien und zielführender Individualurteile in spezifischen Situationen bzw. in einzelnen Trials erhöhen. Bei anderen, nicht dynamischen Aufgabentypen wurde gezeigt, dass die Gruppenleistung bei hoher Demonstrierbarkeit der richtigen Lösung häufig über der mittleren Leistung der Gruppenmitglieder liegt (z.B. Bonner, Sillito, & Baumann, 2007; Henry, 1993; Sniezek & Henry, 1989, 1999). Nur selten übertreffen Gruppen hingegen das Niveau des besten Gruppenmitglieds (Laughlin, Hatch, Silver, & Boh, 2006). In Experiment 1 wurde nachgewiesen, dass die Realgruppenmitglieder auch bei der Festlegung der gemeinsamen Steuerungsurteile bestrebt sind, das subjektiv beste Gruppenmitglied stärker zu gewichten. Die höhere Demonstrierbarkeit eines komplexeren Szenarios sollte es erleichtern, überzufällig gut das beste Gruppenmitglied zu identifizieren und dadurch die Realgruppenleistung zu steigern. Entsprechend ist anzunehmen, dass Realgruppen in Experiment 2 Prozessgewinne realisieren, die über das Leistungsniveau von Nominalgruppen mit gemitteltem und individuellem Systemverlauf hinausgehen. Die im ersten Experiment getesteten Hypothesen 1a und 1b gelten daher entsprechend für Experiment 2.

Trotz der bisherigen Argumentation werden die zwei Alternativhypothesen zum Auftreten (H2a) bzw. Nicht-Auftreten (H2b) von Koordinationsgewinnen aufrechterhalten. Zum einen ist anzunehmen, dass die mit höherer Systemkomplexität und mehr beobachtbaren Variablen einhergehende bessere Demonstrierbarkeit korrekten Wissens und guter individueller Steuerungsvorschläge die Gruppenleistung über das Gruppenpotenzial hinaus verbessert und Koordinationsgewinne auftreten. Zum anderen könnte eine hohe Systemkomplexität aber auch generell die Schwierigkeit der Systemsteuerung erhöhen. Wie Serman (1994) nach Betrachtung zahlreicher Befunde zur dynamischen Systemsteuerungsleistung feststellt, können Menschen kognitiv nur einfache

Kausalstrukturen repräsentieren. Je komplexer ein Steuerungssystem ausfällt, desto stärker weicht die subjektiv erfasste Variablenstruktur von der realen Systembeschaffenheit ab und es finden stärkere Fehlwahrnehmungen der Systemrückmeldungen statt, indem nonlineare Beziehungen oder Nebeneffekte der Steuerungseingriffe ignoriert werden. Kampmann und Serman (1998) konnten in einer dynamischen Wirtschaftssimulation mit Optimierungsziel zeigen, dass die individuelle Steuerungsleistung sich mit zunehmender Systemkomplexität systematisch verschlechterte und mehr inkonsistente Steuerungsurteile auftraten. Für die Steuerung durch Realgruppen könnte man daraus ableiten, dass bei hoher Systemkomplexität mehrere unvollständige und fehlerhafte Systemrepräsentationen aufeinandertreffen, sodass es für die Probanden bei begrenzter Steuerungsdauer eine wesentlich komplexere Aufgabe ist, einen Überblick über antizipierte Variablenzusammenhänge zu gewinnen und Steuerungsurteile in der zweiten Phase festzulegen. Unterschiedliche Vorstellungen innerhalb der Realgruppe könnten auch dazu führen, dass die Gruppe zwischen den Strategien der drei Mitglieder „springt“, um das Zutreffen der individuellen kognitiven Systemrepräsentationen zu überprüfen. Letzteres würde eine konsistente Gewichtung des True Best Member verhindern und somit gegen Koordinationsgewinne aufgrund von G-G Transfer sprechen. Auch das Erkennen der akkuratesten Einzelurteile pro Steuerungstrial könnte aufgrund von bounded rationality (vgl. Simon, 1982), also limitierten Aufmerksamkeits-, Gedächtnis- und Informationsverarbeitungsressourcen angesichts der Fülle an Systeminformationen und Repräsentationen der drei Gruppenmitglieder misslingen. Somit würde auch kein GI-G Transfer in den Realgruppen auftreten und die H2a, die annimmt, dass Gruppen bei der Systemsteuerung nicht ihr Potenzial übertreffen, würde auch für das Experiment 2 bestätigt werden. Die Hypothesen 3 zum individuellen Fertigkeitserwerb und 4 zum Wissenserwerb werden ebenfalls beibehalten und entsprechend der Vorgehensweise in Experiment 1 für das zweite Experiment getestet.

## 5.2 Stichprobe und Design

An der zweiten Untersuchung nahmen 315 Studierende unterschiedlicher Fachrichtungen der Georg-August-Universität Göttingen teil. 187 der Probanden waren weiblich (59.40%) und das Durchschnittsalter der Stichprobe lag bei 22.89 Jahren ( $SD = 2.76$  Jahre). Die Probanden wurden über eine Onlinedatenbank, in der sie sich freiwillig registriert hatten oder über Aushänge im zentralen Hörsaalgebäude der Universität Göttingen rekrutiert. Studierende, die bereits an Experiment 1 mit dem Steuerungsszenario des Kühlsystems in einem Atomkraftwerk teilgenommen hatten, waren von der Teilnahme ausgeschlossen. Experiment 2 beinhaltet ein 3 x 2 Design mit dem dreistufigen

Zwischensubjektfaktor *Bedingung* (Realgruppen- vs. Nominalgruppen- vs. Individualbedingung) und dem zweistufigen Innersubjektfaktor *Steuerungsphase* (Steuerungsphase 1 vs. Steuerungsphase 2). Für einige Analysen wurde zusätzlich der Innersubjektfaktor *Gruppenmitglied* betrachtet.

## 5.3 Methode

### 5.3.1 Steuerungsaufgabe und Leistungsmaß

Die dynamische Systemsteuerungsaufgabe im zweiten Experiment wurde nach folgenden Kriterien ausgewählt: Erstens sollte sie eine höhere Komplexität mit mehr beobachtbaren Variablen aufweisen als das System in Experiment 1. Zweitens sollte die semantische Einkleidung realistisch und für die Studierendenstichprobe nachvollziehbar sein. Drittens sollte das Szenario eindeutig durch ein Gleichungssystem beschrieben sein und die Steuerungsleistung anhand einer quantitativen Outputvariablen des Systems messbar sein. Viertens sollte die Aufgabe bereits zur Untersuchung der individuellen Steuerungsleistung eingesetzt worden sein und dabei relativ stabile, interindividuelle Fertigungsunterschiede aufgezeigt haben.<sup>43</sup>

Die gewählte Aufgabe bestand darin, in der Rolle eines Managers die Produktion und Lagerhaltung eines Unternehmens über 40 Zeittakte so zu steuern, dass die kumulativen Kosten für diese Bereiche minimal gehalten wurden. Das Wirtschaftsszenario wurde in ähnlicher Form bereits von Diehl (1992) zur Untersuchung der individuellen Systemsteuerungsleistung eingesetzt (vgl. auch Diehl & Serman, 1995). Die Systemsimulation wurde, wie das erste Experiment auch, mit der Experimentalsoftware ALEX programmiert (Schlemmer, 2009). Die Steuerungsoberfläche wurde den Probanden computergestützt dargeboten.

Die konkrete Aufgabenstellung mit dem Steuerungsziel wurde den Probanden schriftlich am Laptop vor der ersten Steuerungsphase (und noch nicht vor der Explorationsphase) dargeboten: „Ihre Zielvorgabe als Manager/in ist es, die kumulativen Gesamtkosten über die Gesamtzeit der Systemsteuerung von 40 Durchgängen zu minimieren. Sie können die entstehenden Kosten durch Eingabewerte bei der ‚Änderung der Produktionsrate‘ beeinflussen.“

---

<sup>43</sup> Diese zwei Voraussetzungen sind erforderlich, um Lernprozesse im Gruppenkontext untersuchen zu können. Zum einen darf die Steuerungsleistung nicht nur zufallsbedingt sein, zum anderen sollte das individuelle Leistungslevel relativ stabil sein, sodass die schlechteren Gruppenmitglieder von den besseren lernen können und die Möglichkeit für das Auftreten von G-G Transfer besteht.

***Dauerhaft vorliegende Informationen über Systemparameter und mögliche Parameterzusammenhänge.*** Bereits vor der Explorationsphase wurden die Probanden aufgefordert einen dreiseitigen, in Papierform vorliegenden Text über grundlegende Informationen zur Systemsteuerungsaufgabe „Produktions- und Lagermanagement“ zu lesen. Die Datei ist in Anhang I zu finden. Die Erläuterungen zum Aufgabenszenario begannen mit einer allgemeinen Einleitung und der schematischen Darstellung der auf dem Monitor präsentierten Ein- und Ausgabeoberfläche, die in Abbildung 10 wiedergegeben ist.

„Sie sind Manager/in in einem Produktionsunternehmen und für die Bereiche Produktion und Lagerhaltung zuständig. Genereller Unternehmenszweck ist es, Güter zu produzieren und so die vorhandene Nachfrage zu befriedigen (also so viel zu produzieren, wie auch verkauft werden kann). In jeder Periode (Woche) der Systemsteuerung müssen Sie entscheiden, um wie viele Einheiten Sie die Produktionsrate erhöhen bzw. senken wollen. Die Produktionsrate stellt die Anzahl der in Auftrag gegebenen zu produzierenden Einheiten dar. Als Manager erhalten Sie für jede Urteilsperiode nach Festlegung der aktuellen Produktionsrate eine Rückmeldung über wichtige Variablenwerte des Produktions- und Warenlagersystems in tabellarischer Form (unten ohne Angabe konkreter Werte abgebildet). Die Ausgabe der Variablenwerte erfolgt immer gerundet auf ganze Stückzahlen bzw. auf ganze Euro-Werte.“ Auf den folgenden zwei Seiten des schriftlichen Instruktionmaterials wurden den Probanden nacheinander die Variablenblöcke der Tabelle (ein Block ist jeweils durch eine graue Lehrzeile von den anderen getrennt) erläutert. Hierbei wurden nicht nur tatsächliche Sachverhalte genannt, sondern auch (teils alternativ) mögliche Zusammenhänge zwischen den Systemparametern erläutert. Diese Information sollte ein grundlegendes Verständnis für die relevanten wirtschaftlichen Zusammenhänge bei allen Probanden, unabhängig vom betriebswirtschaftlichen Vorwissen, sicherstellen. Der genaue Wortlaut ist dem Versuchsmaterial I im Anhang zu entnehmen. Die Probanden konnten die Informationsblätter während des gesamten Experimentes einsehen und begannen nach dem erstmaligen Lesen mit der Systemexploration.

Woche	1	2	3	4	5
Änderung der Produktionsrate	_____	_____	_____	_____	_____
Aktuelle Produktionsrate					
<b>Produzierte Einheiten</b>					
Verkaufte Einheiten (abhängig)					
Verkaufte Einheiten (unabhängig)					
<b>Verkaufte Einheiten gesamt</b>					
Veränderung Lagerbestand					
<b>Lagerbestand</b>					
Produktionsanpassungskosten					
Lagerkosten					
<b>Gesamtkosten</b>					
<b>Kumulierte Gesamtkosten</b>					

Abbildung 10. Schematische Darstellung der auf dem Computerbildschirm angezeigten Ein- und Ausgabeoberfläche zur Systemsteuerung. Dargestellt ist der Ausgangszustand vor Eingabe des ersten Wertes für die Änderung der Produktionsrate.<sup>44</sup>

**Zu Grunde liegende Systemgleichungen.** Die im Informationstext beschriebenen oder angedeuteten Zusammenhänge zwischen den Parametern im Produktions- und Lagermanagementszenario sind mathematisch eindeutig definiert und können formal über Systemgleichungen dargestellt werden.<sup>45</sup> Die graphische Darstellung des

<sup>44</sup> Die tabellarische Gestaltung ist angelehnt an die Darstellung Fig. 2.3. Computer Screen Display in Diehl, 1992, S. 72).

<sup>45</sup> Die dargestellten Gleichungen zur Beschreibung des Systems wurden aus Diehl & Sterman (1995) übernommen. Die in den Gleichungen enthaltenen Notationen sind Abkürzungen der englischen Parameterbezeichnungen. Die englischen Begriffe werden in den Erläuterungen jeweils in Klammern genannt.

Produktionsprozesses und der relevanten Systemvariablen in Abbildung 12 dient der Veranschaulichung der Zusammenhänge.

Der zu kontrollierende Lagerbestand (engl. Inventory, I) zu einem Zeitpunkt t (bzw. in einem Trial) ergibt sich aus dem vorhergehenden Lagerbestand ( $I_{t-1}$ ) plus eingehender produzierter Einheiten (Production,  $P_t$ ) minus der verkauften Einheiten (Sales, S).

$$(3) I_t = I_{t-1} + P_t - S_t.$$

Die produzierten Einheiten werden durch die Anzahl veranlasster Produktionseinheiten zum Zeitpunkt  $t-\lambda$  (Production starts,  $P^*_{t-\lambda}$ ) bestimmt. Die Verzögerung zwischen Produktionsveranlassung und Fertigstellung entspricht der Dauer des Produktionsprozesses und wurde auf  $\lambda = 2$  Zeitperioden (in der Tabelle als „Wochen“ bezeichnet) gesetzt.<sup>46</sup>

$$(4) P_t = P^*_{t-\lambda}.$$

Im Experiment bestimmen die Probanden in jedem Trial die Änderung der Produktionsrate  $\Delta_t$ . Die Anzahl veranlasster Produktionseinheiten zum Zeitpunkt t ergibt sich damit aus der vorausgehenden veranlassten Produktionsrate plus der Anpassung um  $\Delta_t$  Einheiten. Die Produktionsrate kann durch positive Werte von  $\Delta_t$  erhöht und durch negative Werte gesenkt werden. Eine Senkung ist maximal auf eine Produktionsrate von 0 Einheiten möglich. Produktionsratenänderungen, die diese natürliche untere Grenze überschreiten, werden automatisch korrigiert.

$$(5) P^*_{t} = P^*_{t-1} + \Delta_t.$$

Produzierte Einheiten fließen in das Lager. Der Lagerbestand wird durch verkaufte Einheiten reduziert (Ein- und Ausgänge aus dem Lager sind in Abbildung 12 durch Pfeile repräsentiert.) Die verkauften Einheiten (Sales, S) setzen sich aus zwei Komponenten, den exogenen Verkäufen X und den endogenen Verkäufen N zusammen.

---

<sup>46</sup> Eine Zeitverzögerung der Systemreaktion um zwei Einheiten wurde gewählt, um diese Systemeigenschaft gegenüber dem ersten Experiment vergleichbar zu halten. Zudem wurde bei der Untersuchung der individuellen Steuerungsleistung bei diesem System festgestellt, dass eine Verzögerung von 2 Einheiten gegenüber 0 und 4 zu einer mittleren Aufgabenschwierigkeit führt (vgl. Diehl, 1992; Diehl & Serman, 1995). Dies begünstigt für die Untersuchung von Gruppenlernprozessen erforderliche, interindividuelle Leistungsunterschiede (gegenüber einer leichten Steuerung ohne Zeitverzögerung, bei der alle Probanden eher gute Leistungen zeigen oder einer schwierigen Steuerung mit 4 Verzögerungseinheiten, die kein Proband mehr gut bewältigen kann).

$$(6) S_t = X_t + N_t.$$

Die exogenen Verkaufszahlen (in Abbildung 10 als „unabhängige Verkäufe“ bezeichnet) repräsentieren den Anteil an der Gesamtnachfrage, der nicht durch das Management des Unternehmens bzw. spezifischer durch Änderungen der Produktionsraten beeinflusst wird. Der Verlauf der exogenen Verkaufszahlen  $U_t$  wurde analog zum Vorgehen von Diehl (1992) durch Generierung von Zufallszahlen aus einer Rechteckverteilung (Uniform Distribution, U) mit einer Range von -15 bis +15 gezogen.<sup>47</sup> Demnach wurde die Anzahl unabhängiger Verkäufe  $X$  zu einem Zeitpunkt  $t$  durch die Anzahl der unabhängigen Verkäufe im vorausgehenden Trial  $X_{t-1}$  plus der Anpassung um die Zufallsschwankung  $U_t$  für diesen Trial berechnet.

$$(7) X_t = X_{t-1} + U_t.$$

Die endogenen Verkäufe  $N$  werden durch Änderungen der Produktionsrate beeinflusst. Es besteht also eine Rückkopplung bzw. ein Nebeneffekt innerhalb des dynamischen Systems, dessen Richtung und Ausprägung die Probanden während der Steuerung erschließen mussten. Die exogene Nachfrage  $N_t$  ergibt sich aus der aktuellen Produktionsrate, multipliziert mit einem Nebeneffektfaktor von  $\gamma = .30$ .

$$(8) N_t = \gamma P_t.$$

Diese positive Rückkopplung oder Verstärkung entspricht realistischen Vorgängen am Markt und wird auch als *Keynesian multiplier* (Diehl & Sterman, 1995, S. 201) bezeichnet. Hierbei führt eine Steigerung der Produktion dazu, dass die Beschäftigung und damit auch die Einkommen steigen. Mit den höheren Einkommen steigt auch die Nachfrage an Gütern, was schließlich die Lager leert und eine weitere Steigerung der Produktion erfordert, um die Nachfrage zu befriedigen.

Die entstehenden Kosten der Systemsteuerung werden sowohl durch den Lagerbestand, als auch durch die Änderung der Produktionsrate bestimmt. Die Formel zur Berechnung der Gesamtkosten ist mit einer Übersicht aller bereits erläuterten, relevanten Systemparameter in Abbildung 11 dargestellt. Lagerkosten entstehen durch (positive oder negative) Abweichungen des Lagerbestandes vom Bestandwert 0. Änderungen der

---

<sup>47</sup> Die Ziehung der Zufallszahlen erfolgte mittels des Statistikprogrammes R (Version 2.11.0). Die 40 Zufallszahlen wurden für die Steuerungsphase 1 in der gezogenen Reihenfolge verwendet und für die Phase 2 in eine neue Reihenfolge gebracht. Somit war die auf der Zufallskomponente basierende Aufgabenschwierigkeit in Bezug auf die eigendynamische Systementwicklung zwischen den beiden Steuerungsphasen identisch, ohne dass Gedächtniseffekte bezüglich des Verlaufes der exogenen Verkaufszahlen die Steuerung in der zweiten Phase erleichtern konnten.

Produktionsrate verursachen ebenfalls Produktionsanpassungskosten, die beispielsweise durch Neueinstellungen bei Produktionserhöhung und durch Entlassungs- und Abfindungskosten bei Produktionssenkung entstehen können. Es wird eine realitätsnahe quadratische Kostenfunktion angenommen. Die Gesamtkosten zu einem Zeitpunkt  $t$  resultieren demnach aus dem Lagerbestand (multipliziert mit dem Faktor  $a = 1$ ) zum Quadrat plus der Änderung der Produktionsrate, die mit dem Faktor  $b = 2$  multipliziert wird, zum Quadrat.<sup>48</sup>

### Systemgleichung zur Kostenermittlung

$$(9) C_t = a * I_t^2 + b * \Delta_t^2.$$

$$= 1 * [I_{t-1} + (P^*_{t-\lambda}) - (X_{t-1} + U_t) + \gamma P_t]^2 + 2 * \Delta_t^2.$$

### Systemparameter

$$a = 1$$

$$b = 2$$

$C_t$  = Gesamtkosten (total costs)

$I_{t-1}$  = Lagerbestand vorher

$P^*_{t-\lambda}$  = Anzahl veranlasster Produktionseinheiten zum Zeitpunkt  $t-\lambda$

$X_{t-1}$  = Exogene Verkäufe vorher

$U_t$  = Änderung der exogenen Verkaufszahl (Zufallskomponente)

$\gamma P_t$  = Endogene Verkäufe

$\Delta_t$  = Änderung der Produktionsrate zum Zeitpunkt  $t$  (Eingabevariable)

$$\lambda = 2$$

$$\gamma = .30$$

Abbildung 11. Dem dynamischen in Experiment 2 zugrunde liegende Kostengleichung und Erläuterungen der Gleichungsparameter im Überblick (Gleichungssystem übernommen aus Diehl & Sterman, 1995).

<sup>48</sup> Nach Diehl und Sterman (1995) haben Simulationsstudien gezeigt, dass die Koeffizienten  $a = 1$  und  $b = 2$  einen guten „trade-off“ (S. 201) zwischen Lager- und Produktionsanpassungskosten darstellen würden und Lösungen zur Kostenminimierung unter diesen Bedingungen zu realistischen Änderungen der Produktionsrate führen würden.



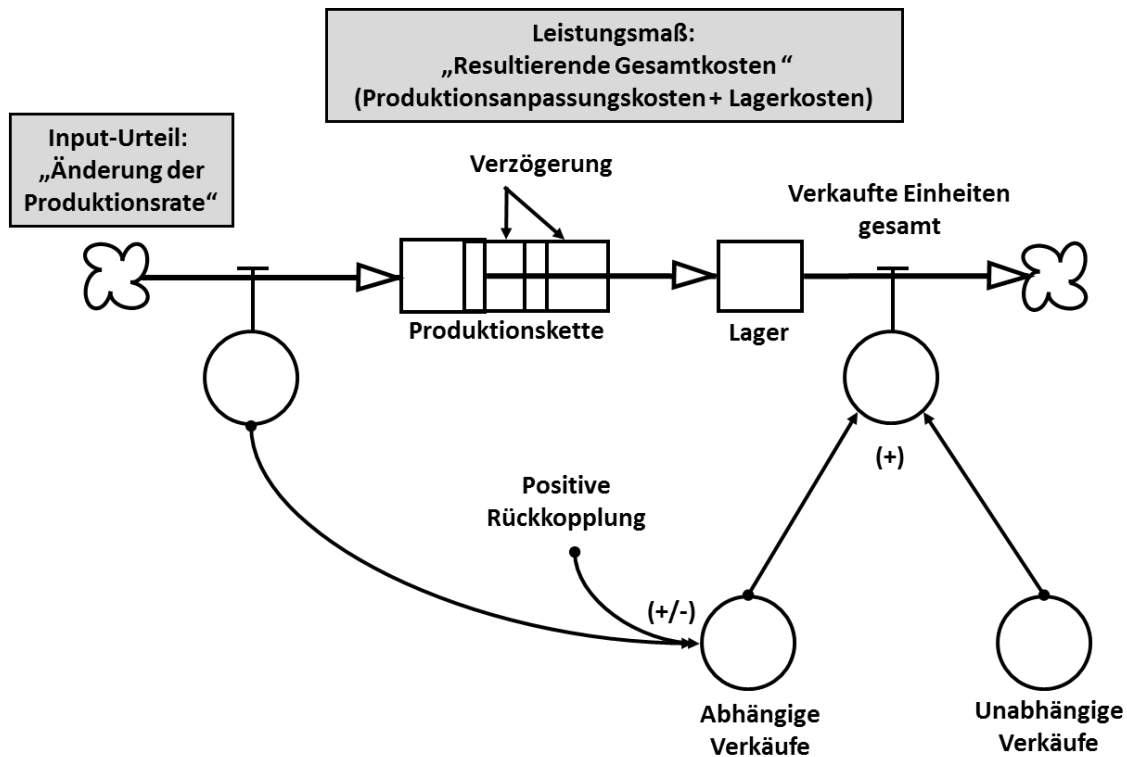


Abbildung 12. Struktur der dynamischen Systemsteuerungsaufgabe in Experiment 2 (In Anlehnung an die Darstellung Fig. 2.1, S.65 in Diehl, 1992).

**Leistungsmaß.** Als abhängige Variable zur Erfassung der Steuerungsleistung wurde der Logarithmus zur Basis 2 der kumulierten Gesamtkosten pro Steuerungsphase ( $\log_2\text{cum\_costs}$ ) berechnet. Die Gesamtkosten pro Trial, die sich aus den Produktionsanpassungskosten und aus den anfallenden Lagerkosten zusammensetzten, wurden über alle 40 Trials einer Steuerungsphase aufsummiert. Bei diesen Gesamtkostenrohwerten ist aufgrund der quadratischen Kostenfunktion eine große Kostenspanne sowie eine linkssteile Verteilung mit hohen Ausreißer- bzw. Extremwerten zu erwarten. Die Gesamtkostenrohwerte wurden (wie auch schon 1992 in der Untersuchung von Diehl) der logarithmischen Transformation zur Basis 2 unterzogen, um eine Annäherung der Verteilung an die Normalverteilung zu erzielen und somit eine parametrische Auswertung der Daten zu ermöglichen.

Je kleiner die Werte des Leistungsmaßes  $\log_2\text{cum\_costs}$  ausfallen, desto besser ist die Steuerungsleistung anzusehen. Zur Beurteilung der absoluten Steuerungsleistung wurden wie im ersten Experiment zwei Benchmarks herangezogen. Die  $\log_2$ -transformierten Kosten nach Anwendung der *optimal rule* stellten die minimal möglichen Kosten bei Lösung des Systemgleichungssystems unter Schätzung der Zufallskomponente

dar. Die genaue Bestimmung dieses Vergleichsmaßstabes einer sehr guten Steuerung wird unter 5.4.1 erläutert. Als zweites Benchmark wurde erneut eine no control rule eingesetzt, bei der über den gesamten Steuerungsverlauf keine Änderungen in der Produktionsrate vorgenommen wurden. Die entsprechende Berechnung des Vergleichsmaßstabes einer schlechten Steuerungsleistung ist ebenfalls in der Ergebnisdarstellung unter Kapitel 5.4.1 beschrieben.

### 5.3.2 Experimenteller Ablauf

Der experimentelle Ablauf verlief weitestgehend identisch zu dem in Abbildung 6 dargestellten Ablaufes des Experimentes 1. Auf die Beantwortung demographischer Angaben folgte eine Systemexplorationsphase mit 20 Eingabetrials. Nach dieser wurde das Steuerungsziel der Minimierung der kumulativen Gesamtkosten bekanntgegeben und alle Probanden durchliefen die individuelle erste Steuerungsphase, wurden auf Basis der individuell erzielten Steuerungsleistung in der ersten Phase zu leistungsheterogen Gruppen zusammengestellt und steuerten dann das System erneut unter den Gegebenheiten der Realbedingung, Nominalbedingung oder Individualbedingung. Nach der Explorations- sowie nach den zwei Steuerungsphasen machten alle Probanden individuelle Angaben zu ihren Annahmen über die Funktionsweise des Systems und ihre Motivation. Der Abschlussfragebogen beinhaltete in der Realbedingung darüber hinaus Fragen zur Kompetenz der Gruppenmitglieder und der subjektiven Gewichtung der Beiträge. Anhang G enthält eine detaillierte Ablaufbeschreibung des Experimentes 2, die die Replikation des Vorgehens ermöglicht. Die mündlichen Instruktionen und Vorgehensweisen der Versuchsleitung finden sich in Anhang H.

Im Gegensatz zu Experiment 1 beinhalteten die Steuerungsphasen jeweils 40 Trials. Die Anzahl der Trials wurde im Vergleich zum ersten Experiment reduziert, um eine annähernd vergleichbare Steuerungsdauer sicherzustellen und Motivationsverlusten aufgrund von Müdigkeit bei zu langer Versuchsdauer vorzubeugen. Die längere Bearbeitungsdauer der Steuerung in Experiment 2 resultierte aus den ausführlicheren Systeminformationen und der größeren zu verarbeitenden Systeminformationsmenge, die die Probanden als Systemoutput nach jeder Steuerungseingabe erhielten. Vor der Explorationsphase lasen die Probanden die unter 5.3.1 erläuterten dauerhaft vorliegenden schriftlichen Informationen zum Produktions- und Lagermanagementsystem (vgl. Anhang I). Für jede Phase standen den Probanden vorstrukturierte Notizenseiten in Papierform mit Eingabefeldern für jeden der 20 bzw. 40 Trials zur Verfügung. Diese sollten eine spätere

Analyse der Vorgehensweise der Probanden ermöglichen.<sup>49</sup> Das Layout der Notizenseiten ist exemplarisch dem Anhang I zu entnehmen. Neu in Experiment 1 war auch ein standardisierter Wissensbogen nach der zweiten Steuerungsphase.

### 5.3.3 Erfassung des Wissenserwerbs

Das verbalisierbare Wissen der Probanden über die Eigenschaften und Funktionsweise des dynamischen Systems wurde mittels zweier Methoden erfasst. Zum einen sollten die Probanden ihre konkreten Annahmen über das System in Form offener Antworten notieren. Zum anderen bearbeiteten die Probanden einen standardisierten Wissenstest mit Multiple-Choice-Antworten, in dem das Verständnis für die zentralen Systemeigenschaften und die Beschaffenheit von Zusammenhängen zwischen den Variablen abgefragt wurden.

**Aus den offenen Antworten resultierende Maße für das Systemwissen.** Die Äußerung von Annahmen im offenen Antwortformat erfolgte jeweils nach der Explorationsphase sowie nach der ersten und nach der zweiten Steuerungsphase. Die Instruktion beinhaltete die vollständige Anweisung aus dem ersten Experiment. Sie wurde allerdings nach der vorangehenden Erfahrung mit der Kodierung der offenen Antworten in Experiment 1 um zwei Aspekte ergänzt.

Die Probanden wurden demnach gefragt, wie ihre konkreten Annahmen über den Zusammenhang zwischen der Änderung der Produktionsrate und den entstehenden Kosten lauten würden. Falls sich ihre Annahmen im Verlauf der Steuerung geändert hätten, sollten sie sie zeitlich so ordnen, dass die aktuell gültige Annahme an letzter Stelle stünde. Im Vergleich zu Experiment 1 folgte die zusätzliche Anweisung, bei mehr als drei Annahmen, alle weiteren mit Nummerierung (4., 5. usw.) in das dritte Textfeld zu schreiben, um die spätere Kodierung zu erleichtern. Außerdem wurden die Probanden explizit aufgefordert, alle Annahmen, die sie während der letzten Steuerungsphase hatten, aufzuschreiben, auch wenn diese während der vorausgehenden Explorations- und/oder Steuerungsphase schon vorhanden gewesen waren! Hierdurch sollten unklare, nicht eindeutig zu kodierende Äußerungen wie „denke immer noch das gleiche“, die sich auf die Annahmen vor der aktuellen Steuerungsphase bezogen, vermieden wären. Solche Kommentierungen waren im ersten Experiment vereinzelt aufgetreten und nicht in die Kodierung einbezogen.

---

<sup>49</sup> Allerdings erfolgten zum Großteil der Trials keine oder nur sehr kryptische Eingaben wie die Notiz nicht genauer benannter Zahlen. Daher wurde von einer Auswertung abgesehen.

Der Inhalt des verbalisierten Wissens wurde entsprechend dem unter 4.3.3 erläuterten Vorgehen in Experiment 1 von zwei unabhängig voneinander arbeitenden Hilfskräften gemäß eines zuvor erstellten kategorialen Auswertungsschemas in einer SPSS-Datendatei kodiert (vgl. für die Kodieranweisung Anhang J). Jede Annahme wurde einer von 16 Inhaltskategorien zugeordnet, die vor der Erhebung des Hauptexperimentes festgelegt worden waren. Die Bezeichnungen der 16 Kategorien sowie Beispielnennungen aus dem Vortestdatensatz und/oder korrespondierende Formulierungen der Multiple-Choice-Antworten aus dem standardisierten Wissenstest zur Veranschaulichung sind in Tabelle 6 aufgelistet.<sup>50</sup> Die Inhaltskategorien mit den Bezeichnungen 6 und 7 wurden als Aufgaben unabhängige mögliche Antworten aus dem ersten Experiment übernommen. Die Kategorie „00“ wurde nach den Probekodierungen der Vortestdaten zur Eignung der Kategorien ergänzt, um entsprechende Annahmen einordnen zu können. Alle anderen Inhaltskategorien wurden aus den zentralen, für die Steuerungsleistung kritischen Eigenschaften des Systems abgeleitet.

Tabelle 6: *Inhaltskategorien zur Kodierung des verbalisierbaren Systemwissens der Probanden in Experiment 2*

Inhalts-kategorie	Kategorienbezeichnung	Beispielnennungen	Richtigkeit der Annahmen
00	Grundannahme: Das System ist nicht stabil/Eigendynamik	Kosten steigen stetig auch bei Nichteingreifen	richtig
0	Optimale Strategie zur Kosten-minimierung richtig:	Produktionsrate möglichst in kleinen Schritten anpassen, Begründung: Anpassungskosten fallen stärker ins Gewicht als Lagerkosten/ Zeitverzögerung einkalkulieren/ Lagerbestand sollte möglichst gering gehalten werden	richtig
1	Zusammenhang Produktionskosten richtig: Erhöhung und Senkung haben vergleichbare Konsequenzen	Größere Veränderungen der Produktionsrate führen zu höheren Produktionsanpassungskosten“ ODER Veränderungen um +/- 4	richtig

<sup>50</sup> Richtige aus dem Wissenstest entnommene Multiple-Choice-Antworten in der Spalte der Beispielnennungen sind mit dem vorangehenden Ausdruck „FB-Item“, der Kurzform für „Fragebogen-Item“ gekennzeichnet. Beispielnennungen ohne diese Markierung wurden aus den Nennungen von Probanden übernommen.

---

		Einheiten verursachen gleich hohe Anpassungskosten ODER <i>FB-Item</i> : Eine Erhöhung der Produktionsrate um x (z.B. 10) Einheiten führte zu <i>genauso großen</i> Produktionsanpassungskosten wie eine Senkung der Produktionsrate um x (z.B. 10) Einheiten	
2	Zeitverzögerte Reaktion des Systems richtig	System reagiert erst später/2 Wochen später	richtig
3	Quadratische Kostenfunktion richtig	Große Änderungen der Produktionsmenge führten zu überproportional hohen Kosten. ODER überproportional hohe Kosten bei größeren Veränderungsschritten	richtig
4	Zusammenhang Produktionskosten falsch	sehr kleine und sehr große Anpassungen der Produktionsrate verursachen höhere Kosten als Anpassungen im mittleren Bereich ODER <i>FB-Item</i> : Eine Erhöhung um x Einheiten führt zu höheren Kosten als eine Senkung um x Einheiten	falsch
5	Zeitverzögerte Reaktion des Systems falsch	System reagiert sofort auf Eingaben	falsch
6	Magisches Zahlendenken	„gerade Zahlen = Kostensteigerung, ungerade Zahlen = Kosteneinsparung	falsch
7	Zufall/kein schlüssiges System	willkürlich?	falsch
8	Zusammenhang Lagerbestand richtig	<b>FB-Item</b> : Ein bestimmter Lagerbestand x (z.B. 20 Einheiten) größer 0 verursachte <i>genauso große</i> Kosten wie ein gleich großer negativer Bestand von x Einheiten, bei dem Sie (z.B. 20) Kunden nicht beliefern konnten.	richtig
9	Funktion Lagerkosten richtig	Große Abweichungen des Lagerbestandes vom Optimum 0 (im positiven oder negativen Bereich) führten zu überproportional hohen Kosten.	richtig

---

10	Kostenverhältnis Produktion und Lagerkosten richtig	<b>FB-Item:</b> Eine Abweichung des Lagerbestandes um x Einheiten (z.B. 10 Einheiten) führte zu <i>halb so großen Kosten</i> wie eine Änderung der Produktionsrate um x (z.B. 10) Einheiten.	richtig
11	Positive Gain bei abhängigen Verkäufen richtig	<b>FB-Item:</b> Eine Erhöhung der Produktion führte zur <b>Steigerung</b> der Anzahl abhängiger Verkäufe.	richtig
12	Zusammenhang Lagerbestand falsch	<b>FB-Item:</b> Ein bestimmter Lagerbestand x (z.B. 20 Einheiten) größer 0 verursachte <i>höhere/niedrigere</i> Kosten als ein gleich großer negativer Bestand von -x Einheiten, bei dem Sie (z.B. 20) Kunden nicht beliefern konnten.	falsch
13	Kostenverhältnis Produktion und Lagerkosten falsch	<b>FB-Item:</b> „Eine Abweichung des Lagerbestandes um x Einheiten (z.B. 10 Einheiten) führte zu <i>genauso großen</i> Kosten wie eine Änderung der Produktionsrate um x (z.B. 10) Einheiten ODER <b>FB-Item:</b> Abweichungen des Lagerbestandes beeinflussten die Gesamtkosten stärker als Veränderungen in der Produktion.	falsch
14	Positive Gain bei abhängigen Verkäufen falsch	<b>FB-Item:</b> Eine Erhöhung der Produktion führte zur <b>Senkung</b> der Anzahl abhängiger Verkäufe.	falsch

*Anmerkung:* Richtige Beispielnennungen, die aus dem Multiple-Choice-Wissenstest entnommen wurden, sind mit dem vorangehenden Ausdruck *FB-Item*, der Kurzform für Fragebogen-Item gekennzeichnet. Beispielnennungen ohne diese Markierung wurden aus den Nennungen von Probanden übernommen.

Gemäß der Erläuterungen zur Berechnung von Kennwerten unter Abschnitt 4.3.3 für das erste Experiment wurden auch für das zweite Experiment zunächst pro Proband jeweils ein Summenwert für die Gesamtzahl aller Annahmen, ein Summenwert für die richtigen Annahmen und ein Summenwert für die falschen Annahmen nach der Explorationsphase, nach der Steuerungsphase 1 und nach der Steuerungsphase 2 berechnet. Anschließend wurden der Prozentsatz richtiger sowie der Prozentsatz falscher Annahmen bestimmt. Aus

den getrennten Berechnungen für die beiden Kodierer wurden dann pro Proband und Explorations- bzw. Steuerungsphase Mittelwerte der resultierenden Gesamtsumme, des Prozentsatzes richtiger und des Prozentsatzes falscher Angaben berechnet. Schließlich wurden die 9 über die Rater gemittelten Kennwerte auf Gruppenebene gemittelt, um als geeignete Maße für das verbalisierbare Systemwissen für die statistischen Berechnungen zur Überprüfung der Hypothesen 4a bis 4c genutzt zu werden.

***Aus dem standardisierten Wissenstest resultierende Maße für das Systemwissen.***

Das zweite Maß für das nach der zweiten Steuerungsphase vorhandene Systemwissen war der standardisierte Wissensfragebogen, der 6 Multiple-Choice-Fragen umfasste. Die Fragen bezogen sich auf Systemeigenschaften und Variablenzusammenhänge, für die im schriftlich vorliegenden Informationsmaterial jeweils mehrere möglicherweise auf das System zutreffende, alternative Zusammenhangsmuster erläutert worden waren. Jede Frage hatte 2 bis 6 Antwortalternativen, von denen jeweils nur eine richtig war.

Für jeden Probanden wurde die Summe der richtigen sowie der falschen Antworten berechnet und anschließend der Prozentsatz richtiger bzw. falscher Antworten bestimmt, indem der Summenwert durch sechs geteilt und anschließend mit 100 multipliziert wurde. Schließlich erfolgte für die Auswertung auf Gruppenebene eine Mittelung der Summen- sowie der Prozentsatzwerte über die drei einer (Nominal-) Gruppe zugeordneten Probanden. Die aus dem Wissenstest berechneten Maße für das Systemwissen waren somit die pro Gruppe gemittelte Gesamtzahl an Annahmen, die gemittelte Anzahl richtiger Antworten bezüglich des Systems bzw. der pro Gruppe gemittelte Prozentsatz richtiger Annahmen über das System.

## **5.4 Ergebnisse und Diskussion**

**Vortest.** Für die Systemsteuerungsaufgabe „Produktions- und Lagermanagement“ wurde ein Vortest mit 32 Probanden durchgeführt, in dem überprüft wurde, ob stabile Fähigkeitsunterschiede bei individueller Bearbeitung auftreten. Ein gewisses Maß an Fähigkeitsstabilität bei der individuellen Steuerung sowie eine gewisse Spannbreite der individuellen Steuerungsleistung sind Voraussetzungen dafür, dass im Gruppenkontext Potenzial besteht, um von den besseren Gruppenmitgliedern zu lernen. Alle Vortestteilnehmer steuerten das System zweimal hintereinander individuell über 40 Takte, wobei die Voreinstellungen des Systems bis auf den Verlauf der Zufallskomponente „exogene Verkäufe“ (siehe oben) identisch war. Der Vortest zeigte eine signifikante positive Korrelation zwischen den zwei Leistungsmaßen für die erste ( $\log_2\_cum\_costs\_ph1$ ) und die zweite ( $\log_2\_cum\_costs\_ph2$ ) Steuerungsphase ( $r_s = .82$ ,  $p = .007$ ). Somit ist die

Voraussetzung der individuellen Leistungsstabilität für den Einsatz der Aufgabe erfüllt. Auch eine hinreichende Leistungsvariabilität war mit einer zur Basis 2 logarithmisierten Kostenspanne von 18.35 bis 37.01 gegeben.

**Überprüfung von Störvariablen in der Hauptuntersuchung.** Für die 315 Probanden der Hauptuntersuchung wurde zunächst der mögliche Einfluss von Alter, Anzahl der Studiensemester, Studienfach und Geschlecht auf die individuelle Steuerungsleistung überprüft. Als Maß für die Steuerungsleistung diente der  $\log_2\_cum\_costs$ -Wert (vgl. 5.3.1) über die Trials 2 bis 40 einer Steuerungsphase. Das Alter korrelierte weder mit der individuellen Steuerungsleistung in Phase 1,  $r(315) = -.03$ ,  $p = .553$ , noch mit der Steuerungsleistung in der Individualbedingung in Steuerungsphase 2,  $r(105) = -.09$ ,  $p = .367$ , jeweils. Das Ausmaß an Studierenerfahrung, gemessen über die Anzahl der bisherigen Studiensemester hatte ebenfalls keinen signifikanten Einfluss auf die Steuerungsleistung in den Phasen 1 und 2,  $r(315) = -.08$ ,  $p = .177$  bzw.  $r(105) = -.17$ ,  $p = .082$ .

Wie in Experiment 1 wurden die Studienfächer in die 5 Kategorien Naturwissenschaften, Geisteswissenschaften, Psychologie, Wirtschaftswissenschaften und eine Restkategorie<sup>51</sup> umkodiert. Die individuelle Steuerungsleistung in Phase 1 unterschied sich signifikant nach Studienrichtung,  $F(4,309) = 4.00$ ,  $p = .004$ ,  $\eta^2_p = .05$ . Einzelvergleiche zeigten, dass Studierende der naturwissenschaftlichen Fächer signifikant bessere Steuerer waren als Geisteswissenschaftler ( $p = .046$ ), Psychologen ( $p = .005$ ) und Studierende mit Fächern der Restkategorie ( $p = .022$ ). Den Studierenden der Wirtschaftswissenschaften waren die Naturwissenschaftler deskriptiv überlegen ( $p = .072$ ). Die Gleichverteilung der überlegenen Studierenden mit naturwissenschaftlichem Fach auf die Bedingungen wurde anhand eines  $\chi^2$ -Tests mit den zwei Kategorien Naturwissenschaftlicher Studiengang *ja* vs. *nein* überprüft. *Es lag kein signifikanter Unterschied zwischen den drei Bedingungen vor*,  $\chi^2(2, N = 314) = 5.30$  (zweiseitig),  $p = .071$ ,  $\phi = .13$ .<sup>52</sup>

---

<sup>51</sup> In die Restkategorie fielen beispielsweise Sportwissenschaften, Geographie oder fehlende Angaben.

<sup>52</sup> Deskriptiv war die Realbedingung mit 11.4% Studierenden der Naturwissenschaften gegenüber der Individualbedingung mit 21.2% sowie der Nominalbedingung mit 22.9% benachteiligt. Die Verteilung der studienfachbezogenen Kompetenz arbeitete also gegen die Hypothese 1 der leistungsbezogenen Überlegenheit von Realgruppen. Die Startkompetenz der Bedingungen, gemessen über die durchschnittliche Individualleistung in Phase 1 unterschied sich jedoch mit  $p = .421$  nicht signifikant.



Geschlechtsunterschiede bei der individuellen Steuerungsleistung in Phase 1 wurden mittels  $t$ -Tests überprüft. Die Männer erreichten signifikant geringere log2-transformierte Gesamtkosten ( $M = 22.05$ ,  $SD = 3.29$ ) und somit eine bessere Steuerungsleistung als die Frauen ( $M = 24.44$ ,  $SD = 3.44$ ),  $t(312) = 6.16$ ,  $p < .001$ ,  $d = 0.71$ . In der Steuerungsphase 2 war die Leistung der 66 weiblichen Probanden in der Individualbedingung ( $M = 23.26$  ( $SD = 4.14$ )) ebenfalls signifikant schlechter als die Leistung der männlichen ( $M = 20.84$ ,  $SD = 2.42$ ),  $t(103) = 3.40$ ,  $p = .001$ ,  $d = 0.71$ . Für die folgenden Analysen waren die Geschlechtsunterschiede nicht problematisch, da keine signifikanten Unterschiede der Geschlechtsverteilung zwischen den Bedingungen vorlagen,  $\chi^2(2, N = 315) = 1.13$ ,  $p = .568$ ,  $\phi = .06$ .

### 5.4.1 Gruppenleistung

**Vergleich mit Leistungsbenchmarks.** Vor der Testung der Hypothese 1 zur Überlegenheit der Realgruppenleistung gegenüber der Leistung in der Nominal- und der Individualbedingung, wurde die absolute Höhe der erzielten Gruppenleistung betrachtet. Hierzu wurden zwei plausible Vergleichsmaßstäbe herangezogen, die die Grenzen der Leistungsspanne von *schlechter* Steuerung zu *guter* Steuerung markieren. Analog zum Vorgehen in Experiment 1 wurde als Vergleichsmaßstab einer eher schlechten Steuerungsleistung der log2\_cum\_costs-Wert berechnet, der ohne Eingriff in das Produktionssystem über die 40 Trials der Steuerungsphasen resultiert. Ohne Veränderungen der anfangs vorhandenen Produktionsrate fallen keine Produktionsanpassungskosten an, und die aus der *no control rule* kumulierenden Gesamtkosten sind lediglich auf anfallende Lagerkosten zurückzuführen und werden maßgeblich von der Zufallskomponente der exogenen Verkäufe bestimmt (vgl. Diehl, 1992, S.99). Unter dem Verlauf der Zufallskomponente in der Steuerungsphase 2 resultierten Gesamtkosten von 38,833,613 Euro. Dies entspricht einem transformierten Wert von  $\log2\_cum\_costs\_ph2_{nocontrol} = 25.21$ .

Als Vergleichsmaßstab einer äußerst *guten* Steuerungsleistung wurde der log2\_transformierte Gesamtkostenwert herangezogen, der bei optimaler Gewichtung aller Systemvariablen zur Herstellung eines Systemgleichgewichts führt (vgl. Diehl, 1922, S. 90 ff.). Unter der *optimal rule* erfolgt die Anpassung der Produktionsrate so, dass ein Lagerbestand von 0 Einheiten angestrebt wird und die Produktion den Verkäufen entspricht. Damit werden die erwarteten Kosten minimiert. Auch unter der *optimal rule* ist die Zufallskomponente des Systems nicht im Vorfeld bekannt und als bestmögliche Schätzung

der zukünftigen unabhängigen Verkäufe wird der aktuelle Wert angenommen.<sup>53</sup> Entsprechend der von Diehl (1992) generierten Simulationsdaten und der Optimalwertbestimmung<sup>54</sup> nach der Ricatti-Gleichung (vgl. D’Azzo & Houpis, 1982) ergibt sich für das hier verwendete System mit einem Nebeneffektfaktor von .30, einer Produktionsdauer von 2 Zeiteinheiten und 40 Steuerungstrials der Optimalwert  $\log_2\_cum\_costs\_ph2_{optimum} = 17.74$ .

Zur Beurteilung der Steuerungsleistung in den Bedingungen wurden jeweils für die Real-, die Nominal- und die Individualbedingung zwei Einstichproben-*t*-Tests gegen die Benchmarks der optimal rule und der no control rule berechnet. Aufgrund der theoretischen Möglichkeit, die Benchmarks der Produktionssteuerung zu übertreffen bzw. zu überbieten, wurde zweiseitig getestet. Tabelle 7 zeigt deskriptive Statistiken der erzielten  $\log_2$ -transformierten Gesamtkosten in den Bedingungen.

Tabelle 7: Mittelwerte, Standardabweichungen und Konfidenzintervalle der mittleren  $\log_2$ -transformierten Gesamtkosten in der Steuerungsphase des Produktions- und Lagermanagementsystems in Experiment 2

Bedingung	<i>n</i>	<i>M</i> ( <i>SD</i> )	95% CI
Real	35	20.48 (2.52)	[19.62, 21.34]
Nominal	35	21.29 (2.10)	[20.57, 22.01]
Individual	35	24.82 (3.20)	[23.72, 25.92]

Anm.: CI = confidence interval (Konfidenzintervall)

Die transformierten erzielten Gesamtkosten fielen in allen Bedingungen signifikant höher aus als der Leistungswert 17.74 unter der optimal rule, alle  $t(34) > 6.44$ , alle  $p < .001$ ,

<sup>53</sup> Da die Vorhersagegenauigkeit der Zufallskomponente durch diese Stabilitätsheuristik von Trial zu Trial schwankt, variiert auch das Ausmaß der Kostenminimierung bei den einzelnen Trials. Wenn die Zufallskomponente nicht korrekt vorhergesagt wird, ist es für die Probanden möglich, die Kostenentstehung unter der optimal rule in einzelnen Trials zu unterbieten. Über die Gesamtsteuerung hinweg ist die optimal rule jedoch nicht zu schlagen.

<sup>54</sup> Diehl führte eine Simulationsstudie mit 225 simulierten Probanden durch und bestimmte anhand der Daten post hoc die optimalen Gewichtungen der Systemvariablen unter der no control rule. Hierbei ergaben sich für das hier verwendete Szenario folgende Gewichte der Inputfaktoren: Lagerbestand = - .3956, Verkäufe = 2.1788, aktuelle Produzierte Einheiten zum Zeitpunkt 1 = 0, Produzierte Einheiten<sub>t+1</sub> = - .3956, Produzierte Einheiten<sub>t+2</sub> = - 1.7832. Diese Gewichtungen sind unabhängig vom konkreten Verlauf des Random Walks, solange er aus der Rechteckverteilung bzw. Uniform Distribution mit einer Range von -15 bis +15 gezogen wird.

alle  $d > 1.09$ . Die Leistung der Realgruppen lag mit  $d = 1.09$  deskriptiv näher am Kostenminimum als die Leistung der Nominalgruppen mit  $d = 1.69$  und die der Probanden in der Individualbedingung mit  $d = 2.62$ .

Der Vergleich mit dem no control rule Benchmark 25.21 zeigte, dass die Realgruppenleistung signifikant besser ausfiel,  $t(34) = -11.12$ ,  $p < .001$ ,  $d = 1.88$ . Auch die Nominalbedingung war der schlechten Steuerungsleistung ohne Kontrolleingriffe überlegen,  $t(34) = -11.01$ ,  $p < .001$ ,  $d = 1.68$ . Die Probanden in der Individualbedingung erreichten bei der Produktionssteuerung hingegen nur ein Leistungsniveau, das weitgehend dem der no control rule entsprach,  $t(34) = -0.72$ ,  $p = .477$ ,  $d = .14$ . Daraus folgt, dass die Steuerungsleistung von Realgruppen und Nominalgruppen mit gemitteltem Systemverlauf als leistungsförderlich im Vergleich zu einem Nichteingreifen bewertet werden kann.

Die Leistung von Nominalgruppen mit individuellem Systemverlauf lag hingegen auf einem relativ schlechten Leistungsniveau, das auch bei Nichteingreifen in das Produktionssystem resultiert hätte. In der Real- und der Nominalbedingung überschritten nur jeweils 2 Gruppen mit einer schlechten Leistung das no control rule Benchmark. In der Individualbedingung waren es 11 der 35 post hoc gebildeten Gruppen, die schlechtere Steuerungsergebnisse erzielten als die no control rule.<sup>55</sup>

Abbildung 13 veranschaulicht den Leistungsverlauf unter den drei Bedingungen über die pro Trial verursachten (untransformierten) Gesamtkosten. Während die Systemsteuerung in der Nominal- und der Realbedingung über den gesamten Verlauf der Phase 2 nicht mehr als Kosten in Höhe von 2 Millionen pro Trial verursacht, kommt es in der Individualbedingung zu einer deutlichen Destabilisierung des Produktionssystems und stark ansteigenden Kosten in der zweiten Hälfte der Steuerungsphase 2.

---

<sup>55</sup> Häufigkeitsunterschiede zwischen den Bedingungen wurden nicht statistisch ausgewertet, da der Anteil der erwarteten Häufigkeiten, die kleiner als 5 waren, jeweils 20% überschritt und somit die Voraussetzung zur Durchführung eines  $\chi^2$ -Tests nicht erfüllt war (vgl. Bortz, 1999, S. 170). Keine der Gruppen in der Real- oder Individualbedingung erreichte das Leistungsniveau der Optimal Rule. In der Nominalbedingung gelang dies einer Gruppe.

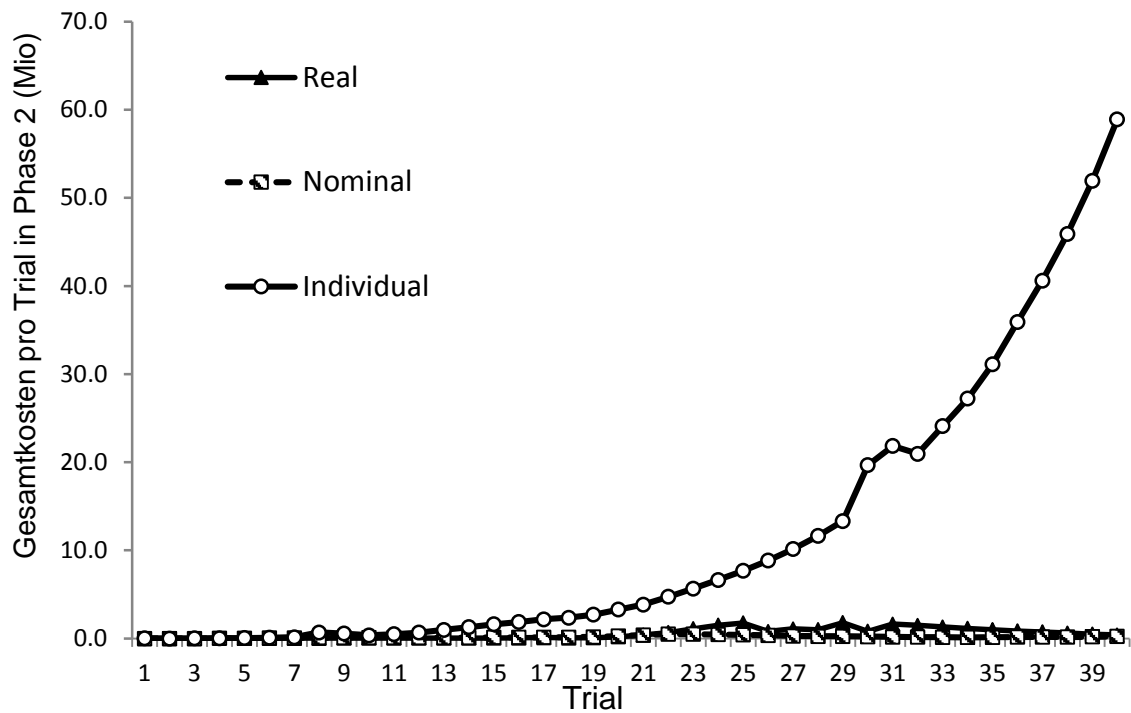


Abbildung 13. Mittlere erzielte Gesamtkosten pro Bedingung für jeden der 40 Trials der Steuerungsphase 2.

**Untersuchung von Prozessgewinnen.** Ob interagierende Gruppen Prozessgewinne realisierten und eine bessere Leistung bei der Steuerung des dynamischen Systems erbrachten als Nominalgruppen mit gemitteltem Systemverlauf oder als Nominalgruppen mit individuellem Systemverlauf wurde mit einer ANOVA untersucht. Das Leistungsmaß als abhängige Variable wurde über die Differenz aus der Startkompetenz der Gruppen und der Leistung in der zweiten Gruppenphase berechnet, um Unterschiede im Ausgangsniveau der Gruppen herauszurechnen. Die Startkompetenz bestand aus den pro Gruppe gemittelten  $\log_2$ -transformierten Kosten bei der individuellen Produktionssteuerung in Phase 1. Die Leistung in der zweiten Phase entsprach in der Realbedingung den  $\log_2$ -transformierten Kosten, die aus den Gruppenurteilen für die 40 Trials der Steuerungsphase 2 resultierten. In der Nominalbedingung resultierten die  $\log_2$ -transformierten Kosten aus den vom Servercomputer generierten Gruppenurteilen über die 40 Trials der Phase 2. In der Individualbedingung wurden jeweils drei durch individuelle Produktionssteuerung erzielte  $\log_2$ -transformierte Gesamtkostenwerte gemittelt, um einen Nominalgruppenleistungswerte bei individuellem Systemverlauf zu berechnen. Da niedrigere Gesamtkosten einer besseren Steuerungsleistung entsprachen, bedeuteten

positive Werte des Differenzmaßes zwischen Startkompetenz und Gruppenphasenleistung eine Leistungszunahme in Phase 2.

Die ANOVA zeigte einen signifikanten Effekt für die Bedingung  $F(2,102) = 36.38$ ,  $p < .001$ ,  $\eta^2_p = .42$ . Bonferroni-adjustierte Kontraste ergaben, dass die Realgruppenleistung ( $M = 3.25$ ,  $SD = 2.34$ ) und die Nominalgruppenleistung ( $M = 2.07$ ,  $SD = 2.10$ ) jeweils signifikant besser ausfielen als die Leistung in der Individualbedingung ( $M = -1.439$ ,  $SD = 3.102$ ), beide  $p < .001$ ,  $d = 1.71$  bzw.  $d = 1.33$ . Daher wird die Hypothese 1a, die besagt, dass die Systemsteuerungsleistung real interagierender Gruppen der Steuerungsleistung von Nominalgruppen mit individuellem Systemverlauf in der Individualbedingung überlegen ist, angenommen. Das Differenzmaß der Realgruppenleistung unterschied sich zwar deskriptiv hypothesenkonform zur Hypothese 1b von der Nominalgruppenleistung, der Unterschied zwischen den beiden Bedingungen war aber bei zweiseitiger Testung nicht signifikant,  $p = .126$ ,  $d = 0.53$ . Somit wird die Hypothese 1b nicht angenommen, obwohl eine Effektstärke mittlerer Größe vorlag, die auf einen systematischen Unterschied hindeuten könnte.<sup>56</sup> Bei größerer Stichprobe und höherer Testpower ist davon auszugehen, dass man eine signifikante Überlegenheit der Realgruppen über den statistischen Aggregationsvorteil der Nominalgruppen mit gemitteltem Verlauf zeigen kann. Wahrscheinlich hätte sich der Leistungsunterschied zwischen der Real- und der Nominalbedingung ebenfalls stärker manifestiert, wenn die Steuerungsdauer auf mehr als 40 Trials ausgeweitet gewesen wäre.

#### 5.4.2 Lerneffekte auf Gruppenebene

Im Folgenden wurde analysiert, ob die Gruppen der Realbedingung während der Gruppeninteraktion Koordinationsgewinne erzielten. Koordinationsgewinne liegen im Produktionsmanagementszenario vor, wenn die aus den Gruppenurteilen resultierenden Kosten geringer ausfallen als die aus dem Gruppenpotenzial theoretisch resultierenden Kosten. Wie im ersten Experiment wurde das Gruppenpotenzial post hoc über den Mittelwert der Individualurteile pro Trial der Steuerungsphase 2 berechnet. Da die

---

<sup>56</sup> Wie in Experiment 1 wurde explorativ berechnet, ob die Steuerungsleistung zwischen den drei Bedingungen für die zweite Hälfte der Phase 2 (hier mit den log<sub>2</sub>-transformierten Gesamtkosten der Trials 21 bis 40 einen deutlicheren Leistungsvorteil der Realbedingung anzeigte. Das Ergebnismuster entsprach aber auch in Experiment 2 der Analyse der Gesamtphase,  $F(2,102) = 25.21$ ,  $p < .001$ ,  $\eta^2_p = .33$ . Die Leistungen in der Real- ( $M = 19.72$ ,  $SD = 2.77$ ) und Nominalbedingung ( $M = 20.32$ ,  $SD = 2.59$ ) fielen beide besser aus als die Leistung in der Individualbedingung ( $M = 24.36$ ,  $SD = 3.49$ ), beide  $p < .001$ . Sie unterschieden sich aber erneut nicht signifikant voneinander,  $p = .700$ .

Individualurteile im aI-G Design jeweils pro Trial vor den Gruppenurteilen abgegeben wurden, waren möglicherweise auftretende individuelle Fertigkeitsergebnisse (I-I und/oder G-I Transfer) aus diesem Maß herausgerechnet. In weiteren Einzelanalysen wurde jeweils die zu erwartende Leistung unter Zutreffen des Accuracy, des Initial Best Member sowie des True Best Member Models mit der Realgruppenleistung verglichen. Schließlich wurde mit Hilfe von RMSD-Werten untersucht, welches der normativen Modelle die Steuerungsleistung der Realgruppen am besten vorhersagte.

**Gruppenleistung vs. Average Model.** Zunächst wurde überprüft, ob die Realgruppenleistung bei der Produktionssteuerung das Gruppenpotenzial übertraf und Koordinationsgewinne auftraten. Hierzu wurde post hoc der Mittelwert aus den individuellen Gruppenmitgliederurteilen pro Trial der Steuerungsphase 2 bestimmt und mit SPSS die gemäß des Systemgleichungssystems theoretisch resultierenden Gesamtkosten berechnet. Diese Gesamtkosten des Average Models sowie die tatsächlich verursachten Kosten durch Realgruppensteuerung wurden zur Basis 2 logarithmisiert und in einem *t*-Test für abhängige Stichproben miteinander verglichen. Die Gruppenleistung ( $M = 20.48$ ,  $SD = 2.52$ ) fiel signifikant schlechter aus, als es das Gruppenpotenzial bzw. das Average Model ( $M = 20.38$ ,  $SD = 2.50$ ) ermöglicht hätte,  $t(34) = 6.81$ ,  $p < .001$ ,  $d = 1.15$ . Der *t*-Test zeigte somit, dass die Realgruppen bei der Produktionssteuerung keine Koordinationsgewinne erzielten, sondern signifikant höhere Kosten verursachten, als bei einer Mittelungsstrategie der Individualurteile resultiert wären.<sup>57</sup> Somit wird die Hypothese 2a angenommen. Eine mögliche Erklärung für dieses Ergebnis ist, dass die Realgruppen einer differenziellen Gewichtungstrategie gefolgt sind und das beste Gruppenmitglied oder die kostengünstigsten Individualurteile stärker gewichtet haben. Falls die Individualurteile sehr stark um das optimale Urteil streuen und keinen systematischen Bias<sup>58</sup> aufweisen, resultiert die Auswahl des besten Urteils in einer größeren Abweichung (und schlechteren Leistung) als der Mittelwert der Individualurteile.<sup>59</sup> In den folgenden Analysen wurde überprüft, ob die

---

<sup>57</sup> Ein Wilcoxon-Vorzeichen-Rang-Test für verbundene Stichproben mit den nicht-transformierten Gesamtkosten in der Steuerungsphase 2 als abhängiger Variable zeigte analog, dass Gruppen eine signifikant schlechtere Leistung erbrachten, als es das Gruppenpotenzial ermöglicht hätte,  $z = 5.01$ ,  $p < .001$ , zweiseitige Testung: 32 der 35 Realgruppen verursachten größere Kosten als das Average Model.

<sup>58</sup> Unter Bias wird eine systematische Urteilsverzerrung der Individualurteile im Sinne einer mittleren Über- oder Unterschätzung des optimalen Steuerungsurteils innerhalb der Gruppen verstanden.

<sup>59</sup> Für den Fall, dass die Gruppenmitglieder das beste Individualurteil bzw. das beste Gruppenmitglied nicht zu 100% gewichten, sondern eine moderat stärkere Gewichtung gemäß

Gruppenleistung mit dem Leistungslevel des Accuracy Models, des Initial Best Member Models oder des True Best Member vergleichbar war.

**Root Mean Squared Difference (RMSD): Passung der Vergleichsmodelle.** Wie auch für das Experiment 1 wurde abschließend die relative Passung der normativen Modelle des Average Models, des Accuracy Models, des True Best Member Models und des Initial Best Member Models auf die Gruppenleistung über den RMSD bestimmt. Hierzu wurde die Wurzel aus den mittleren quadrierten Abweichungen der unter den Vergleichsmodellen resultierenden Gesamtkosten von den unter den Realgruppenurteilen resultierenden Gesamtkosten der Trials 1 bis 40 der zweiten Steuerungsphase berechnet. Je kleiner der RMSD ausfällt, desto besser entspricht ein Vergleichsmodell der Realgruppenleistung.

Die Berechnung der Kostenwerte nach dem Average Model wurde oben bereits erläutert. Die Steuerungsurteile des Accuracy Models entsprachen pro Steuerungstrial dem Individualurteil innerhalb der Realgruppe, das die geringsten Gesamtkosten pro Trial nach sich zog. Der normative Vergleichswert des Initial Best Member Models bestand in der Steuerungsleistung, die aus der vollständigen Gewichtung des in Phase 1 individuell besten Gruppenmitglieds in allen Trials der Gruppenphase resultiert hätte. Als vierter normativer Vergleichsmaßstab für die Realgruppenleistung wurde das True Best Member Model berechnet. Hierbei wurde das aufgrund der Individualurteile in Phase 2 insgesamt beste Realgruppenmitglied ausgewählt. Aus dessen Individualurteilen wurden post hoc die theoretisch resultierenden Gesamtkosten der Produktionssteuerung berechnet. Die RMSD-Werte betragen 326,754.28 für das Average Model, 331,517.59 für das Accuracy Model, 504,291.25 für das Best Member Model und 465,308.23 für das True Best Member Model.

In separaten Wilcoxon-Tests zeigte sich, dass der RMSD des Average Models signifikant niedriger ausfiel als die RMSD-Werte des Initial Best Member Models und des Best Member Models, beide  $Z > 3.51$ , beide  $p < .001$ . Der RMSD des Accuracy Models war mit  $z = 1.47$  und  $p = .140$  nicht signifikant verschieden vom RMSD des Average Models. In 24 der 35 Gruppen wies das Average Model eine bessere Passung auf die Realgruppendaten auf, in 11 Fällen das Accuracy Model. Dieses Ergebnis zeigt, dass die Mehrheit der Realgruppen während der zweiten Steuerungsphase annähernd einer

---

Urteilsqualität bzw. Expertise vornehmen, sollte sich die differenzielle Gewichtung hingegen positiv auf die Gruppenleistung in Relation zum Gruppenpotenzial auswirken.

Mittelungsstrategie folgte, aber 11 der Gruppen die Gruppenurteile stärker am besten Individualurteil pro Trial ausrichteten.

Das Average Model zeigte somit deskriptiv die beste Passung, um die Steuerungsergebnisse der Realgruppen zu beschreiben. Die Realgruppenleistung fiel jedoch noch schlechter aus, als unter hundertprozentiger Mittelung der Individualurteile zu erwarten gewesen wäre. Damit wird die Alternativhypothese 2a angenommen, die davon ausgeht, dass keine Koordinationsgewinne auftreten und die auf den Gruppenurteilen basierende Steuerungsleistung vergleichbar oder schlechter ausfällt als die aus dem Mittelwert der individuellen Urteile der Gruppenmitglieder resultierende Steuerungsleistung. Bei der Steuerung des Produktions- und Lagermanagementsystems blieben Koordinationsgewinne nicht nur aus, sondern es traten sogar signifikante Koordinationsverluste auf. Die mittlere absolute (nicht log<sub>2</sub>-transformierte) Kostendifferenz zwischen den durch die Realgruppensteuerung verursachten Gesamtkosten und den theoretisch aus dem Gruppenpotenzial hervorgehenden Gesamtkosten betrug -2.77 Mio € bzw. Kosteneinheiten ( $SD = 1.43$ ). Die Spanne der Kostendifferenz für die einzelnen Realgruppen reichte von Koordinationsverlusten in Höhe von 84.76 Mio € bis zu Gewinnen in Höhe von 0.01 Mio €. Die erhöhte Systemkomplexität mit vielen beobachtbaren Variablen scheint sich somit nicht (durch erhöhte Demonstrierbarkeit von Systemwissen und leistungsförderlichen Strategien) positiv auf die Festlegung der Gruppensteuerungsurteile auszuwirken. Stattdessen ist es plausibel, dass die Informationsfülle der Systeminformationen und die unterschiedlichen geäußerten Systemrepräsentationen und Strategievorschläge der Realgruppenmitglieder die kognitiven Verarbeitungskapazitäten überstiegen und zu einer dysfunktionalen Koordination der individuellen Steuerungsvorschläge der Gruppenmitglieder führten. Hinweise darauf, dass eine erhöhte Informationsmenge die Realgruppenleistung (in Relation zum Gruppenpotenzial) negativ beeinflusst haben könnte, finden sich in Studien zum leistungsmindernden Effekt erhöhter Informationsfülle auf die individuelle Systemsteuerungsleistung (Kampmann & Sterman, 1998) sowie auf die Urteilsleistung bei Schätzaufgaben und geringer individueller Ausgangsfähigkeit (Bonner & Baumann, 2008).

### 5.4.3 Lerneffekte auf individueller Ebene

Im Folgenden wurde untersucht, ob die Probanden bei der wiederholten individuellen Systemsteuerung in den Phasen 1 und 2 Leistungssteigerungen erzielten. In der Individualbedingung entspricht eine Verbesserung der Steuerungsleistung in der zweiten Phase I-I Transfer bzw. Übungseffekten. In der Real- und der Nominalbedingung



kann eine Leistungsverbesserung in Bezug auf die Individualurteile während der Gruppenphase drei mögliche Ursachen haben, die gleichzeitig bzw. anteilig wirksam sein können (vgl. die Ergebnisdiskussion in Kapitel 4.4.6). Erstens kann I-I Transfer vorliegen. Zweitens kann G-I Transfer stattfinden, indem Gruppenmitglieder ihre Individualurteile an den Urteilen oder Strategien besserer Gruppenmitglieder in der Realbedingung ausrichten bzw. die generierten mittleren Urteile in der Nominalbedingung als Referenzwert nutzen. Drittens werden die Individualurteile in Phase 2 bei Ausgangszuständen des Systems getroffen, die auf den Gruppenurteilen bzw. den generierten gemittelten Individualurteilen basieren. Durch die statistische Aggregation entstehen so auch günstigere Ausgangszustände für eine gute individuelle Steuerungsleistung als in Phase 1. Der Anteil von I-I Transfer an der Individualleistung in der Real- bzw. Nominalbedingung kann über die Berechnung in der Individualbedingung geschätzt werden. Für die individuelle, über den I-I Transfer hinausgehende Leistungssteigerung kann mit dem Design des Experimentes 2 nicht differenziert werden, zu welchen Anteilen dabei G-I Transfer und/oder ein statistischer Aggregationsvorteil wirksam werden.

Die individuelle Leistungsveränderung wurde für jeden Probanden über die Differenz aus den individuellen log<sub>2</sub>-transformierten Gesamtkosten der ersten Steuerungsphase und den aus den Steuerungsurteilen resultierenden log<sub>2</sub>-transformierten Gesamtkosten der zweiten Steuerungsphase berechnet.<sup>60</sup> Positive Werte des Differenzmaßes indizieren eine individuelle Leistungssteigerung von der ersten zur zweiten Steuerungsphase. Aufgrund der Abhängigkeit der Daten auf Gruppenebene in der Real- und Nominalbedingung durch denselben Systemverlauf und dieselbe Interaktionserfahrung bzw. Referenzwertinformation, erfolgte eine Auswertung auf Gruppenebene. Hierzu wurde der Innersubjektfaktor *Gruppenmitglied* in die Analyse aufgenommen wurde. Eine 3 (*Bedingung*: Realbedingung vs. Nominalbedingung vs. Individualbedingung) x 3 (*Gruppenmitglied*: bestes vs. mittleres vs. schlechtestes) ANCOVA mit der Bedingung als Zwischensubjektfaktor, dem Gruppenmitglied als Innersubjektfaktor und der Startkompetenz der Gruppen als Kovariate berechnet. Die Zuordnung der Gruppenmitglieder als bestes, mittleres und schlechtestes Gruppenmitglied ergab sich aus der relativen individuellen Leistung in der ersten Steuerungsphase. Die Startkompetenz beschreibt das mittlere individuelle Leistungsniveau der Gruppe in der individuellen Steuerungsphase 1 und wurde einbezogen, um die Fehlervarianz zu reduzieren. Je

---

<sup>60</sup> Die aus den individuellen Urteilen in der Real- und in der Nominalbedingung hervorgehenden Temperaturen wurden nachträglich mit SPSS berechnet.

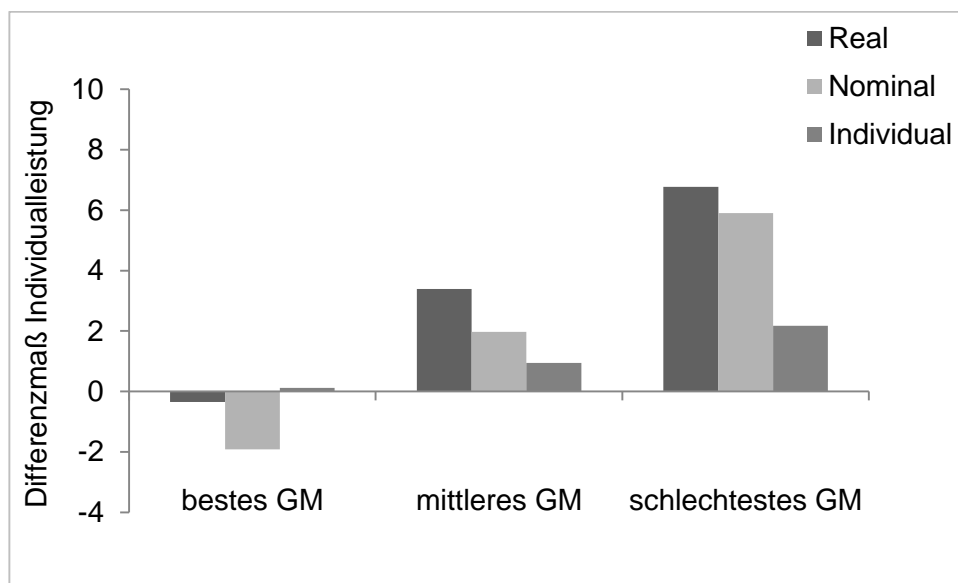
niedriger das durchschnittliche Leistungslevel einer Gruppe vor der Gruppeninteraktion war, desto mehr Spielraum zur individuellen Leistungsverbesserung bestand.

In der ANCOVA lag ein signifikanter Effekt der Bedingung vor,  $F(2, 101) = 10.47$ ,  $p < .001$ ,  $\eta^2_p = .17$ . Bonferroni-korrigierte post-hoc Tests zeigten, dass die individuelle Leistungssteigerung in der Realbedingung unter Berücksichtigung der Startkompetenz mit  $p < .001$  signifikant größer ausfiel als die Übungseffekte in der Individualbedingung, und mit  $p = .034$  ebenfalls größer ausfiel als in der Nominalbedingung mit gemitteltem Systemverlauf. Der entsprechende Leistungszuwachs zwischen der Nominalbedingung und der Individualbedingung unterschied sich hingegen mit  $p = .145$  nicht signifikant. Zusätzlich berechnete post-hoc Einzel- $t$ -Tests gegen Null zeigten, dass sich die individuellen log2-transformierten Gesamtkosten in allen Bedingungen signifikant verringerten und ein bedeutsamer (von Null verschiedener) Leistungszuwachs vorlag. Die Leistungssteigerung in der Realbedingung ( $M = 3.27$ ,  $SD = 2.28$ ),  $t(34) = 8.48$ ,  $p < .001$ ,  $d = 1.43$ , fiel dabei größer aus als die Verbesserung in der Nominalbedingung ( $M = 1.98$ ,  $SD = 2.04$ ),  $t(34) = 5.74$ ,  $p < .001$ ,  $d = .97$ . Die Leistungsverbesserung durch I-I Transfer war im Vergleich zu den anderen zwei Bedingungen am geringsten, aber immer noch signifikant von Null verschieden ( $M = 1.08$ ,  $SD = 1.36$ ),  $t(34) = 4.67$ ,  $p < .001$ ,  $d = .79$ .

Damit kann die Hypothese 3a angenommen werden, die besagt, dass die Verbesserung der individuellen Steuerungsleistung von Gruppenmitgliedern in der Realgruppenbedingung größer ist als die Verbesserung der Steuerungsleistung von Nominalgruppenmitgliedern mit individuellem Systemverlauf. Auch die Hypothese 3b wird durch die Leistungsdaten bei der dynamischen Produktionssteuerung gestützt, da die Verbesserung der individuellen Steuerungsleistung von Gruppenmitgliedern in der Realgruppenbedingung größer ausfällt als die Verbesserung der Steuerungsleistung von Nominalgruppenmitgliedern mit gemitteltem Systemverlauf.

Der Haupteffekt für die Kovariate Startkompetenz war mit  $F(1, 101) = 3.84$ ,  $p = .053$ ,  $\eta^2_p = .04$  knapp nicht signifikant. Dennoch erreichten die Gruppenmitglieder in Gruppen mit einem durchschnittlich schlechteren Ausgangsniveau wie erwartet im Mittel größere individuelle Leistungszuwächse,  $r(105) = .22$ ,  $p = .023$ . Die Interaktion aus Gruppenmitglied und Startkompetenz war über die drei Bedingungen nicht signifikant,  $F(1.73, 175.13) = 2.13$ ,  $p = .129$ ,  $\eta^2_p = .021$ . Der Haupteffekt für den Faktor Gruppenmitglied war ebenfalls nicht signifikant,  $F(1.73, 175.13) = .407$ ,  $p = .637$ ,  $\eta^2_p = .00$ . Wie auch beim Kühlsystemszenario des ersten Experimentes lag im Produktionsmanagementszenario ein signifikanter Interaktionseffekt von Gruppenmitglied und Bedingung vor,  $F(3.47, 175.13) = 24.63$ ,

$p < .001$ ,  $\eta^2_p = .33$ . Abbildung 14 zeigt die individuellen Leistungsdifferenzen zwischen Phase 1 und Phase 2 für das jeweils beste, mittlere und schlechteste Gruppenmitglied pro Bedingung. Deskriptiv entspricht das Ergebnismuster dem des Experimentes 1: Die Unterschiede im individuellen Fertigkeitsszuwachs zwischen den Mitgliedern in der Individualbedingung waren schwächer ausgeprägt als in der Real- und in der Nominalbedingung. Detailliertere Analysen anhand von Einzel- $t$ -Tests für das jeweils beste, mittlere und schlechteste Gruppenmitglied pro Bedingung sind in Anhang K verfügbar.



*Abbildung 14.* Ausmaß der individuellen Leistungssteigerung für das gemäß der Individualleistung in Phase 1 beste, mittlere und schlechteste Gruppenmitglied (GM) in Experiment 2. G-I Transfer wurde über ein Differenzmaß aus den log2-transformierten Gesamtkosten der Phase 1 und der Phase 2 bestimmt.

#### 5.4.4 Wissenserwerb bei der Systemsteuerung

Der Wissenserwerb der Probanden im Verlauf des Experimentes wurde mit zwei Methoden erhoben: Erstens über selbst verbalisierte schriftliche Angaben der Probanden und zweitens über einen Wissenstest mit Multiple Choice Fragen. Das verbalisierbare Wissen der Probanden über die Zusammenhänge im Produktions- und Lagermanagementsystem wurde zu drei Zeitpunkten (jeweils nach der Explorationsphase, sowie nach der ersten und nach der zweiten Steuerungsphase) über das Schreiben der individuellen Annahmen in offene Textfelder erfasst. Wie unter Punkt 4.3.3 erläutert,

beurteilten zwei unabhängige Kodierer die Richtigkeit der Annahmen und ordneten sie je einer der 14 Inhaltskategorien zu. Pro Steuerungsphase und Proband wurde nach der Kodierung jeweils die Gesamtsumme an Annahmen, die Summe richtiger Annahmen sowie der Prozentsatz richtiger Annahmen berechnet. Die Interraterreliabilität für die Summenscores zur Kodierung *richtig vs. falsch* betrug  $Kappa = .76$  ( $SE = .01$ ),  $p < .001$ , 95% CI [.74, .78]. Dies ist als substantielle Übereinstimmung (*substantial agreement*) zu bewerten (Landis & Koch, 1977). Die Interraterreliabilität für die inhaltliche Kodierung (die daraufhin den Kategorien richtig vs. falsch zugeordnet und pro Proband aufsummiert wurde) betrug  $Kappa = .67$  ( $SE = .01$ ),  $p < .001$ , CI [.66, .68].

Aufgrund der gegebenen Interraterreliabilität wurden die beiden Kodierungen für jeden Probanden gemittelt, und zwar in Bezug auf die Gesamtsumme an Annahmen, die Anzahl richtiger Annahmen und den Prozentsatz richtiger Annahmen über das Wirtschaftssystem.<sup>61</sup> Dies geschah wie in Experiment 1 auch für die Explorationsphase, die Steuerungsphase 1 und die Steuerungsphase 2. Anschließend wurden die Werte der Probanden auf Gruppenebene gemittelt, um der Abhängigkeit der Daten gerecht zu werden. Mit diesen Kennwerten wurden die Hypothesen 4a bis 4c überprüft, die davon ausgehen, dass Realgruppenmitglieder durch die Interaktion einen quantitativen und qualitativen Wissensvorteil gegenüber individuell oder in Nominalgruppen mit gemitteltem Systemverlauf steuernden Probanden erlangen. Entsprechend der Teilhypothesen 4a bis 4c wurde dabei die *relative* Gesamtzahl an Annahmen, die Anzahl richtiger Annahmen und der Prozentsatz richtiger Annahmen zwischen den Bedingungen verglichen.

Für die statistische Analyse der Hypothesen zum relativen Wissenserwerb in den drei Bedingungen wurde pro Proband jeweils ein Differenzmaß aus der ersten und zweiten Steuerungsphase für jede der drei abhängigen Variablen (Gesamtzahl Annahmen, Anzahl richtiger Annahmen und Prozentsatz richtiger Annahmen) berechnet. Anschließend wurde der Mittelwert für jede Gruppe berechnet. Das Differenzmaß berücksichtigt Unterschiede im Ausgangswissen nach der ersten individuellen Steuerungsphase und ermöglicht Aussagen darüber, inwieweit sich die Quantität bzw. die Korrektheit des Wissens durch die Gruppeninteraktion im Vergleich zur Nominal- und Individualbedingung verändert hat. In Tabelle 8 sind die Mittelwerte und Standardabweichungen für die Differenzmaßwerte pro Bedingung zu finden. Links davon finden sich die zugehörigen Systemwissenswerte, die nach der zweiten Steuerungsphase erfasst wurden.

---

<sup>61</sup> Grundlage war hierbei die inhaltliche Kodierung der beiden Kodierer, die dann für die Auswertung nach den übergeordneten Kategorien richtig vs. falsch ausgewertet wurde.

Tabelle 8: Mittelwerte und Standardabweichungen der drei abhängigen Variablen Gesamtzahl der Annahmen, Anzahl richtiger Annahmen und Prozentsatz richtiger Annahmen für die absoluten Werte in Steuerungsphase 2 sowie für das Differenzmaß aus Phase 1 und Phase 2 in Experiment 2

Variable	<i>n</i>	Werte Phase 2 <i>M</i> ( <i>SD</i> )	<i>n</i>	Differenzmaß <i>M</i> ( <i>SD</i> )
Gesamtzahl Annahmen				
Realbedingung	35	1.59 (.70)	35	-.10 (1.38)
Nominalbedingung	35	1.46 (.71)	35	-.41 (1.20)
Individualbedingung	35	1.70 (.60)	35	-.40 (1.21)
Anzahl richtig				
Realbedingung	35	1.08 (.48)	35	-.09 (1.25)
Nominalbedingung	35	.86 (.55)	35	-.25 (.74)
Individualbedingung	35	1.10 (.58)	35	-.34 (1.18)
Prozentsatz richtig				
Realbedingung	34	70.15 (24.24)	24	-.59 (34.21)
Nominalbedingung	34	58.15 (26.11)	27	8.88 (39.12)
Individualbedingung	35	62.21 (21.75)	30	-5.06 (35.40)

*Anmerkung:* Die unvollständigen Zellbesetzungen bei der Variablen Prozentsatz richtig resultierten daraus, dass in den fehlenden Fällen keine Annahmen nach der Steuerungsphase 2 getroffen wurden.

Zur Testung der Hypothesen 4a bis 4c wurde jeweils eine ANOVA mit dem Zwischensubjektfaktor Bedingung (Realbedingung vs. Nominalbedingung vs. Individualbedingung) und dem Hypothesen spezifischen Differenzmaß als abhängiger Variable berechnet. Für keines der drei Differenzmaße der Gesamtzahl der Annahmen,  $F(2,102) = 0.69$ ,  $p = .506$ ,  $\eta^2_p = .01$ , der Anzahl richtiger Annahmen,  $F(2,102) = 0.45$ ,  $p = .626$ ,  $\eta^2_p = .01$ , sowie dem Prozentsatz korrekter Annahmen,  $F(2,78) = 0.78$ ,  $p = .464$ ,  $\eta^2_p = .02$ , lag ein signifikanter Bedingungsunterschied vor. Somit wird sowohl die Hypothese 4a, die besagt, dass Personen in Realgruppen mehr Annahmen über die Systemeigenschaften treffen als Nominalgruppen mit gemitteltem oder individuellem Systemverlauf, als auch die Hypothese 4b, die davon ausgeht, dass die Anzahl korrekter

*Annahmen* über Systemeigenschaften bei Realgruppenmitgliedern signifikant höher ausfallen sollte als bei der Vergleichsbedingungen, sowie die Hypothese 4c, die annimmt, dass Realgruppenmitgliedern den Probanden in den beiden Vergleichsbedingungen im Prozentsatz korrekter Annahmen über die Systemeigenschaften überlegen sind, abgelehnt. Da keine signifikanten Bedingungsunterschiede in Bezug auf das Systemwissen vorlagen, erübrigte sich die in Experiment 1 an dieser Stelle angeschlossene Mediationsanalyse.

Schließlich wurde in Bezug auf die offenen Antworten exploriert, ob die Wissenskenwerte mit der erzielten Steuerungsleistung (operationalisiert über die log2-transformierten kumulativen Kosten in Phase 2) korrelierten. Die Korrelation zwischen den log2-transformierten kumulativen Kosten und der Gesamtsumme der Annahmen in Phase 2 betrug  $r = .09$ ,  $p = .384$ . Zwischen der Anzahl richtiger Annahmen und dem Leistungsmaß,  $r = -.02$ ,  $p = .833$  sowie zwischen dem Prozentsatz richtiger Annahmen und der Leistung bestand ebenfalls kein Zusammenhang,  $r = -.04$ ,  $p = .705$ . Eine gute Steuerungsqualität ging demnach nicht systematisch mit umfangreicherem verbalisierbaren Wissen über die Funktionsweise des Produktions- und Lagermanagementsystems einher.

Als zweites standardisiertes Maß wurde ein Wissenstest mit 6 Multiple-Choice Fragen nach der zweiten Steuerungsphase am PC durchgeführt. Die Fragen bezogen sich auf Richtung und Größe der Auswirkungen von Änderungen der Produktionsrate und unterschiedlichen Abweichungen vom optimalen Lagerbestand im Produktionsszenario. Alle Probanden beantworteten die Fragen individuell. Post hoc wurde pro Proband die Summe der richtigen Antworten und anschließend der Mittelwert der Anzahl korrekter Antworten pro Gruppe berechnet. Zwischen den drei Bedingungen gab es keinen signifikanten Unterschied im Verständnis der zentralen Systemzusammenhänge,  $F(2,102) = 0.546$ ,  $p = .581$ ,  $\eta^2_p = .01$ . Die Realgruppenmitglieder beantworteten im Durchschnitt 2.89 Fragen korrekt ( $SD = .91$ ) und schnitten damit deskriptiv etwas besser ab als die Probanden in der Nominalbedingung ( $M = 2.75$ ,  $SD = .60$ ) und in der Individualbedingung ( $M = 2.71$ ,  $SD = .61$ ). Die Korrelation zwischen der Summe richtiger Antworten im Wissenstest und der Steuerungsleistung war ebenfalls nicht signifikant,  $r = -.18$ ,  $p = .074$ .

#### **5.4.5 Explorative Analysen: Erkennen von steuerungsbezogener Expertise und subjektive Gewichtung im Realgruppenkontext**

Genauso wie im ersten Experiment schätzen die Probanden nach der zweiten Steuerungsphase unabhängig voneinander die eigene Kompetenz und die der anderen

Realgruppenmitglieder ein. Eine relative Kompetenzbeurteilung fand statt, indem alle drei Gruppenmitglieder in eine Rangreihe gebracht wurden. Zusätzlich erfolgte eine absolute Kompetenzeinschätzung jedes Gruppenmitglieds auf einer Skala von 1 (*sehr geringe Kompetenz*) bis 7 (*sehr hohe Kompetenz*). Die Richtigkeit der relativen Kompetenzeinschätzung des pro Gruppe besten bzw. schlechtesten Mitglieds wurde post hoc bestimmt, indem die Rangzuordnung in eine dichotome Variable rekodiert wurde (1 = richtiger Rangplatz, 0 = falscher Rangplatz). Anschließend wurde der Mittelwert pro Gruppe berechnet. Die Identifikation des besten (bzw. schlechtesten) Mitglieds durch alle drei Gruppenmitglieder entsprach somit einem Wert von 1, die richtige Identifikation durch zwei Gruppenmitglieder einem Wert von .67 und die Identifikation durch ein Gruppenmitglied einem Wert von .33. Der Wert 0 resultierte, wenn kein Gruppenmitglied den richtigen Rangplatz zuwies. Anschließend wurde der gruppenbasierte Identifikationswert mit dem Erwartungswert unter zufälliger Zuordnung eines Rangplatzes zu einem Mitglied (.33) verglichen.

Den Realgruppenmitgliedern gelang es in 57% der Fälle und somit überzufällig häufig das initial best member (die Person mit der besten individuellen Steuerungsleistung in Phase 1) zu identifizieren,  $t(34) = 3.90$ ,  $p < .001$ ,  $d = 0.65$ . Ebenso entsprach die Zuordnung des schlechtesten Rangplatzes in 53% der Fälle überzufällig häufig dem initial worst member,  $t(34) = 3.03$ ,  $p = .005$ ,  $d = 0.50$  (vgl. Tabelle 9). Von dieser Kompetenzzuordnung hätten die Realgruppen für die Ausrichtung der Gruppenurteile am besten Mitglied profitieren können, falls die Rangreihe auch während der Gruppensteuerungsphase stabil geblieben wäre. In zwei  $t$ -Tests wurde überprüft, ob das beste und schlechteste Mitglied gemäß der Initialleistung in Phase 1 auch überzufällig häufig das beste bzw. schlechteste Mitglied gemäß der Individualleistung in Phase 2 war. Das beste Gruppenmitglied behielt mit 49% der Fälle nicht überzufällig häufig den Rangplatz bei,  $t(34) = 1.82$ ,  $p = .078$ ,  $d = 0.30$ . Ein Expertisefeedback bezüglich der individuellen Produktionsmanagementleistung in Phase 1 hätte den Realgruppen somit nur in 17 der 35 Fälle das individuell beste Gruppenmitglied während der Gruppenphase aufgezeigt. Auch das anfangs schlechteste Mitglied war mit 37% nicht überzufällig häufig dasjenige mit der schlechtesten Leistung in Phase 2,  $t(34) = 0.50$ ,  $p = .620$ ,  $d = 0.08$ . Entsprechend dieser geringen Fertigkeitstabilität in Bezug auf die Rangreihe in der Gruppe, erkannten die Probanden weder überzufällig häufig das gemäß der Individualleistung während der Gruppenphase beste Mitglied (true best member),  $t(30) = -0.26$ ,  $p = .793$ ,  $d = 0.05$ , noch das individuell schlechteste Gruppenmitglied (true

worst member),  $t(30) = 0.38$ ,  $p = .705$ ,  $d = 0.07^{62}$ . Der prozentuale Anteil richtiger Kompetenzeinschätzungen für das beste und schlechteste Mitglied während der Gruppenphase ist in Tabelle 9 dargestellt.

Tabelle 9: Mittelwerte und Standardabweichungen des Prozentsatzes richtig identifizierter relativer Mitgliederkompetenz in der Realbedingung des Experimentes 2

Bedingung	<i>n</i>	<i>M</i> ( <i>SD</i> )
Initial best member	35	.57 (.37)
Initial worst member	35	.53 (.40)
True best member	35	.31 (.35)
True worst member	35	.30 (.39)

Als zweites Maß der Kompetenzbeurteilung nach der Gruppenphase wurde eine *absolute* Kompetenzeinschätzung für jedes einzelne Gruppenmitglied auf einer Skala von 1 (*sehr geringe Kompetenz*) bis 7 (*sehr hohe Kompetenz*) durchgeführt. Post hoc wurde das pro Gruppe gemittelte Kompetenzrating für das beste, mittlere und schlechteste Gruppenmitglied gemäß der Individualleistung in Phase 2 berechnet. Eine MANOVA mit dem Innersubjektfaktor *Gruppenmitglied* (true best member vs. true medium member vs. true worst member) zeigte keinen Unterschied im mittleren Kompetenzrating,  $F(2,68) = 0.59$ ,  $p = .555$ ,  $\eta^2_p = .02$ . Die mittlere Kompetenzeinschätzung des true best member war mit 4.55 ( $SD = 1.02$ ) deskriptiv sogar geringfügig niedriger als die des true medium ( $M = 4.73$ ,  $SD = .85$ ) und die des true worst member ( $M = 4.81$ ,  $SD = 1.10$ ).<sup>63</sup>

<sup>62</sup> Darüber hinaus war die Gruppenleistung nicht besser, wenn mehr Mitglieder innerhalb einer Gruppe das true best member erkannten,  $r_s(35) = -.07$ ,  $p = .689$ .

<sup>63</sup> Eine entsprechende ANOVA mit dem initial member Status (Zuordnung als bestes, mittleres und schlechtestes Gruppenmitglied auf Basis der Individualleistung in Phase 1) als unabhängiger Variable war hingegen erwartungsgemäß signifikant,  $F(2,68) = 16.68$ ,  $p < .001$ ,  $\eta^2_p = .33$ . Die mittlere Kompetenzbeurteilung des initial best member fiel mit  $p = .270$  ( $M = 5.18$ ,  $SD = .89$ ) deskriptiv höher aus als die Beurteilung des initial medium member ( $M = 4.86$ ,  $SD = .87$ ) und mit  $p > .001$  signifikant höher als die des initial worst member ( $M = 4.06$ ,  $SD = .87$ ). Das initial medium member wurde außerdem als signifikant kompetenter eingestuft als das initial worst member.



Im Folgenden wurde in zwei separaten Analysen exploriert, ob sich die subjektive Gewichtung<sup>64</sup> der Individualurteile bei der Gruppenurteilsfestlegung in der Gruppenphase an der subjektiven (und häufig inkorrekten) Expertiserangreihe orientierte oder ob sie alternativ gemäß der tatsächlichen Leistungsrangreihe erfolgte. Eine MANOVA mit dem Innersubjektfaktor *subjektive Gruppenmitgliedsexpertise* (subjective best vs. subjective medium vs. subjective worst member) zeigte einen signifikanten *subjektiven* Gewichtungsunterschied zwischen den Gruppenmitgliedern,  $F(1.65, 56.01) = 28.48, p < .001, \eta^2_p = .46$ . Die Realgruppenmitglieder gaben an, die Individualurteile des subjective best member ( $M = 38.62, SD = 5.54$ ) und des subjective medium member ( $M = 36.40, SD = 6.77$ ) jeweils signifikant stärker als die des subjective worst member ( $M = 27.16, SD = 6.47$ ) berücksichtigt zu haben, beide  $p < .001$ . Die subjektiven Gewichtungen des subjective best und des subjective medium member unterschied sich mit  $p = .393$  nicht signifikant.

In einer zweiten MANOVA mit dem Innersubjektfaktor *wahre Gruppenmitgliedsexpertise* (true best member vs. true medium member vs. true worst member) und der pro Gruppe gemittelten subjektiven Gewichtung pro Mitglied als abhängiger Variable wurde überprüft, ob die intendierte Gewichtung sich zwischen den nach individueller Steuerungskompetenz in Phase 2 eingeteilten Gruppenmitgliedern unterschied. Dies war nicht der Fall,  $F(2, 68) = 0.50, p = .610, \eta^2_p = .01$ . Die Realgruppenmitglieder gewichteten subjektiv die Individualurteile des true best member ( $M = 32.82, SD = 8.61$ ), des true medium member ( $M = 34.55, SD = 7.12$ ) und des true worst member ( $M = 34.81, SD = 8.19$ ) gleich stark und folgten somit nach eigenen Angaben einer Mittelungsstrategie. Diese Angaben der Probanden stimmen mit der unter 5.4.2 festgestellten Passung des Average Models auf die Gruppendaten überein.

Zusammenfassend kann festgehalten werden, dass die Gruppenmitglieder im Produktionsmanagementszenario (wie schon im Kühlsystemszenario des ersten Experimentes) die Intention hatten, die Individualurteile des besten Gruppenmitglieds stärker als die des schlechtesten für die Gruppenurteile zu gewichten. Die relative Kompetenzzuordnung entsprach überzufällig häufig der Individualleistung der Steuerungsphase 1. Aufgrund der geringen Stabilität der gruppeninternen

---

<sup>64</sup> Die Probanden schätzten nach der Realgruppeninteraktion ein, wie stark die individuellen Urteile pro Gruppenmitglied über alle Trials hinweg in das Gesamturteil der Systemsteuerwerte Ihrer Gruppe eingeflossen seien. Dabei machten Sie prozentuale Angaben, die sich für alle drei Einschätzungen zu 1 bzw. 100% aufaddierten. Als abhängige Variable in der MANOVA wurde die mittlere subjektive Gewichtung aller drei Probanden pro Gruppenmitglied berechnet. Das Gewichtungsmaß basierte also auf einer Selbst- und zwei Fremdeinschätzungen.

Leistungsrangreihe zwischen der ersten und zweiten Steuerungsphase entsprach die Kompetenzeinschätzung nicht mehr überzufällig häufig der relativen Individualleistung während der Gruppenphase. Dies hatte möglicherweise zur Folge, dass die Realgruppenmitglieder die Beiträge des tatsächlich besten Mitgliedes während der Interaktion nicht stärker gewichteten und unbeabsichtigt eine „Mittelungsstrategie“ bei der Festlegung der Gruppenurteile einsetzten.<sup>65</sup>

#### 5.4.6 Zusammenfassung

Das Ziel von Experiment 2 war es, die Replizierbarkeit der Befunde des wenig komplexen Systemszenarios aus Experiment 1 in einem komplexeren dynamischen Szenario zu untersuchen. Hierzu wurde ein computersimuliertes Produktions- und Lagermanagementsystem mit 11 beobachtbaren Systemvariablen (vgl. Diehl, 1992; Diehl & Serman, 1995) und dem zentralen Steuerungsziel der Gesamtkostenminimierung verwendet. Das experimentelle Design und der grundlegende Versuchsablauf aus Experiment 1 wurden beibehalten. Die Probanden steuerten das Wirtschaftssystem in der ersten Steuerungsphase über 40 Trials individuell. In der zweiten Steuerungsphase steuerten sie das System erneut über 40 Trials entweder in der Realbedingung in interagierenden Dreiergruppen, in der Nominalbedingung in räumlich getrennten, computervernetzten Dreiergruppen, deren Individualurteile vom Serverrechner gemittelt wurden, oder in der Individualbedingung wiederholt individuell.

Um das Auftreten von Prozessgewinnen zu untersuchen, wurde ein Differenzmaß aus den auf Gruppenebene gemittelten, log2-transformierten Gesamtkosten der ersten Steuerungsphase mit den entsprechenden Gesamtkosten der zweiten Steuerungsphase berechnet. Das Befundmuster zu Prozessgewinnen des ersten Experimentes zeigte sich auch in Experiment 2: Die Gruppenleistung in der Realbedingung fiel hypothesenkonform signifikant besser aus als die Steuerung der Probanden in der Nominalbedingung. Die Realgruppenleistung war der Leistung von Nominalgruppen mit gemitteltem Systemverlauf

---

<sup>65</sup> Diese Interpretation der unbeabsichtigten Mittelungsstrategie ist zwar plausibel, aber nicht die einzig mögliche Erklärung für das Datenmuster. Alternativ könnte auch eine absichtliche Mittelung der Urteile durch die Probanden stattgefunden haben, und sie könnten im anschließenden Fragebogen eine beabsichtigte differenzielle Gewichtung angegeben haben, um zu den von ihnen zuvor angegebenen wahrgenommenen Kompetenzunterschieden zwischen den Mitgliedern konsistent zu sein.

jedoch nicht signifikant überlegen, obwohl ein Effekt mittlerer Größe bei einer deskriptiv besseren Realgruppenleistung vorlag.

Das Befundmuster zu Lerneffekten auf Gruppenebene wies Gemeinsamkeiten und Unterschiede zu den Ergebnissen aus Experiment 1 auf. Repliziert werden konnte der zentrale Befund, dass Realgruppen bei Vergleich mit dem dynamisch angepassten Gruppenpotenzial im al-G Design keine Koordinationsgewinne erzielten. Allerdings entsprach die Realgruppenleistung in Experiment 2 nicht wie zuvor dem Leistungsniveau des Gruppenpotenzials, sondern es traten signifikante Koordinationsverluste auf. Der Vergleich mit weiteren normativen Modellen zeigte, dass die Gruppenleistung ebenfalls signifikant schlechter ausfiel, als unter dem Accuracy Model, dem Best Member Model oder dem True Best Member Model zu erwarten gewesen wäre. Das Average Model wies jedoch die beste Passung auf die Realgruppendaten auf.

In explorativen Analysen wurde analog zu Experiment 1 untersucht, welche Ursachen es hatte, dass die Realgruppen keine Koordinationsgewinne erzielten. Dabei zeigten sich zwei potenziell leistungsförderliche Sachverhalte: Zum einen hatten die Probanden bei der Steuerung des Produktionsmanagementszenarios die Absicht, das Gruppenurteil stärker an den Individualurteilen des subjektiv besten und des zweitbesten Gruppenmitglieds auszurichten als am subjektiv schlechtesten Gruppenmitglied. Zum anderen gelang es den Probanden auch überzufällig häufig die Kompetenzrangreihe gemäß der Individualleistung in der ersten Steuerungsphase zu erkennen. Der kritische leistungsmindernde Punkt war die geringe Stabilität der gruppeninternen Leistungsrangreihe zwischen der ersten und der zweiten Steuerungsphase. Die Zuordnung des richtigen Kompetenzranges innerhalb der Gruppe für das True Best Member und True Worst Member gemäß der Individualleistung in Phase 2 lag auf Zufallsniveau. Gemäß der wahren Mitgliederkompetenzrangreihe entsprachen die subjektiven Gewichtungsangaben der Realgruppenmitglieder einer „Mittelungsstrategie“. Somit decken sich die Probandenangaben mit dem ermittelten relativ guten Fit des Average Models auf die Gruppenleistungsdaten. Im Vergleich zu Experiment 1 hätte Leistungsfeedback nicht nur für das beste, sondern auch für das schlechteste Gruppenmitglied keine überzufällige Vorhersagevalidität gehabt und stellt daher keine empfehlenswerte Intervention zur Leistungssteigerung dar.

Zur Untersuchung der Teilhypothesen zum individuellen Fertigkeitengewinn in der Steuerungsphase 2 wurde die individuelle Leistungsveränderung über die Differenz aus den individuellen log<sub>2</sub>-transformierten Gesamtkosten der ersten Steuerungsphase und den

aus den Steuerungsurteilen resultierenden log<sub>2</sub>-transformierten Gesamtkosten der zweiten Steuerungsphase berechnet. Die Probanden in der Individualbedingung zeigten signifikante Übungseffekte bzw. I-I Transfer. Anders als in Experiment 1 bestätigten sich beide Teilhypothesen zur individuellen Leistungssteigerung: Die Leistungssteigerung der Realgruppenmitglieder übertraf sowohl signifikant die Übungseffekte in der Individualbedingung als auch die Leistungssteigerung in der Nominalbedingung. Der Anteil am individuellen Leistungszuwachs in der Realbedingung, der über den in der Nominalbedingung hinausging, war mit hoher Wahrscheinlichkeit interaktionsbedingt. Allerdings ermöglicht Experiment 2 keine Aussage darüber, ob die Realgruppenmitglieder durch die Gruppeninteraktion dauerhaft ihre individuelle Systemsteuerungskompetenz steigern konnten, und auch ohne alternierend zu ihren Individualurteilen stattfindende Gruppeninteraktion das individuelle Leistungslevel halten.

Schließlich wurde untersucht, ob ein hypothesenkonformer Wissenszuwachs durch die Realgruppeninteraktion entstand, der über die Steigerung der Wissensquantität und Wissensqualität in der Nominal- und Individualbedingung hinausging. Das Wissen wurde zum einen durch frei formulierte Äußerungen der Probanden zur Funktionsweise des Produktions- und Lagermanagementsystems und zum anderen durch einen Multiple-Choice Fragebogen mit 6 Items erfasst. Die Anzahl richtiger Antworten im Wissensfragebogen lag bei knapp 50% und unterschied sich nicht zwischen den Bedingungen. Wie in Experiment 1 nahm die Quantität des frei formulierten Wissens in allen Bedingungen ab. Die Entwicklung der Anzahl richtiger Annahmen und des Prozentsatzes richtiger Annahmen unterschied sich ebenfalls nicht signifikant zwischen den Bedingungen.

## 6. Experiment 3

### 6.1 Zielsetzung und Hypothesen

Die Experimente 1 und 2 konnten konsistent zeigen, dass face-to-face interagierende Gruppen (unter Berücksichtigung der durchschnittlichen individuellen Ausgangsleistung) eine signifikant bessere Leistung bei der dynamischen Systemsteuerung erbrachten als Nominalgruppen, die sich aus derselben Anzahl individuell arbeitender Probanden zusammensetzten. Ein signifikanter Leistungsvorteil von Realgruppen gegenüber Nominalgruppen mit gemitteltem Systemverlauf konnte bisher nicht gezeigt werden. Es fanden sich aber Hinweise darauf, dass sich der Nutzen der Gruppeninteraktion weniger in der Gruppenleistung als vielmehr im Lernzuwachs der Gruppenmitglieder ausdrückt. Die mittlere individuelle Steuerungsleistung der Realgruppen verbesserte sich von der ersten individuellen Steuerungsphase zur zweiten Gruppensteuerungsphase signifikant stärker als dies durch Übungseffekte bei wiederholter individueller Systemsteuerung in der Individualbedingung der Fall war. Eine mögliche Erklärung für die Überlegenheit der Realgruppen über die durchschnittliche individuelle Steuerung stellt Lerntransfer von der Gruppe auf das Individuum (G-I Transfer) dar. Durch die Übernahme besserer Steuerungsstrategien oder Referenzwerte für die Systemeingaben war es beispielsweise für die relativ schlechten Realgruppenmitglieder möglich von den besseren zu lernen und das individuelle Steuerungsniveau während der Gruppenphase zu steigern. Die Probanden in der Individualbedingung hatten diese Möglichkeit nicht. Meyer und Scholl (2009) liefern einen Hinweis darauf, dass der Austausch in der Gruppe während der gemeinsamen Systemsteuerung zu G-I Transfer führen könnte. In ihrer Studie konnten Meyer und Scholl (2009) in einem computerbasierten Systemszenario einer Schneiderwerkstatt zeigen, dass die individuelle Systemsteuerung davon profitierte, wenn zuvor in Dyaden aus Mitgliedern mit experimentell induziertem, teilweise (nicht) überlappenden Systemwissen über das System diskutiert wurde. Allerdings wurde die individuelle Steuerungsleistung nicht vor und nach der Gruppendiskussion erfasst, sodass bei Meyer und Scholl kein Rückschluss auf G-I Transfer möglich ist. Auch fand in diesem Experiment keine gruppenbasierte Systemsteuerung statt.

Aufgrund der methodischen Vorgehensweise in den Experimenten 1 und 2 können die Befunde aber nur als ein Hinweis auf (statt als Beleg für) G-I Transfer gewertet werden. Grund dafür ist, dass die individuellen Steuerungsurteile in der Realbedingung, aus denen post hoc die theoretisch resultierende individuelle Steuerungsleistung berechnet wurde, alternierend zu den Gruppenurteilen abgegeben wurden, die den Systemverlauf der

zweiten Steuerungsphase bestimmten. Damit profitierte auch die individuelle Steuerung vom statistischen Aggregationsvorteil der Gruppenurteile, die verglichen mit der Individualbedingung relativ günstige Ausgangszustände für eine gute Steuerungsleistung hervorriefen. Der individuelle Fertigkeitserfolg bei den Nominalgruppen mit gemitteltem Systemverlauf war in beiden vorangehenden Experimenten nur deskriptiv geringer als der Gewinn in der Realbedingung und ebenfalls signifikant größer als die Übungseffekte in der Individualbedingung. Für die Nominalbedingung ist offen, ob dieser scheinbar große individuelle Fertigkeitserfolg auch bei individueller Steuerung nach der Nominalgruppenphase erhalten bleibt (da er beispielsweise auf der Kenntnis der Gruppenmittelwertinformationen als gute Referenzwerte basiert), oder ob er ein statistisches Artefakt ist und die unabhängige Individualleistung auf das Niveau der Individualbedingung zurückfällt.

Hierzu wird der experimentelle Ablauf um eine zweite individuelle Steuerungsphase nach der Gruppenphase erweitert. Es liegt somit ein I-G-I-Design vor. Durch die Berechnung der individuellen Leistungsdifferenz aus der ersten und dritten Steuerungsphase kann der individuelle Lernzuwachs bei individuellem Systemverlauf untersucht werden, ohne dass eine Konfundierung mit der „Günstigkeit“ der Systemausgangsbedingungen und dem Gruppenkontext vorliegt. In diesem Design kann zusätzlich für die Nominalbedingung exploriert werden, ob der vermeintliche individuelle Leistungszuwachs während der zweiten Phase in einer dritten individuellen Steuerungsphase bestehen bleibt, oder nur auf einem statistischen Artefakt beruht.

Um eindeutig G-I Transfer nachzuweisen, muss die individuelle Leistung nach der interaktiven Gruppensteuerungsphase (bzw. nach der Nominalgruppenphase mit individuellem oder gemitteltem Systemverlauf) bei unabhängiger Steuerung mit individuellem Systemverlauf gemessen werden. Somit ist eine Entkonfundierung von der Qualität der Systemausgangszustände gewährleistet. Bei einer signifikant größeren Individualleistungssteigerung der Realgruppenmitglieder nach der Gruppeninteraktion gegenüber der individuellen Verbesserung der Nominalgruppenmitglieder der Vergleichsbedingungen kann auf G-I Transfer geschlossen werden. Dieser G-I Transfer wäre nicht nur gegenüber den Übungseffekten der Individualbedingung sondern auch gegenüber dem Lernen von Referenzwerten durch Gruppenmittelwertinformationen in der Nominalbedingung überlegen.

Das Hauptziel des dritten Experimentes ist folglich der Nachweis, dass bei der dynamischen Systemsteuerung in Realgruppen G-I Transfer auftritt, der auch über die

Gruppensituation hinaus stabil ist. Darüber hinaus wird im dritten Experiment die Replizierbarkeit der bisherigen Befunde zur Gruppenleistung, zu individuellen Fertigungsgewinnen in der zweiten Steuerungsphase, zu Koordinationsgewinnen sowie zum Erwerb von Systemwissen überprüft. Um die Ergebnisdarstellung konsistent zu den ersten beiden Experimenten zu halten, wird die bisher verwendete Auswertungsreihenfolge beibehalten und das Vorliegen stabilen G-I Transfers nach der Analyse individueller Fertigungsgewinne in der Steuerungsphase 2 überprüft. Entsprechend wird die Hypothese zu stabilem G-I Transfer durch gruppenbasierte Systemsteuerung im dritten Experiment mit 3 c bezeichnet:

**Hypothese 3 c:** Die Verbesserung der individuellen Steuerungsleistung von Realgruppenmitgliedern nach der Gruppeninteraktion fällt größer aus als die Verbesserung der Steuerungsleistung von Nominalgruppenmitgliedern mit individuellem Systemverlauf und Nominalgruppenmitgliedern mit gemitteltem Systemverlauf.

## 6.2 Stichprobe und Design

315 Studierende unterschiedlicher Fachrichtungen der Georg-August-Universität Göttingen nahmen an der dritten Untersuchung teil. 172 der Probanden waren weiblich (54.60%). Das durchschnittliche Alter der Stichprobe lag bei 23.52 Jahren ( $SD = 3.75$  Jahre). Die Probanden wurden über eine Onlinedatenbank, in der sie sich freiwillig registriert hatten, oder über Aushänge im zentralen Hörsaalgebäude der Universität rekrutiert. Studierende, die bereits an Experiment 1 oder Experiment 2 teilgenommen hatten, waren von der Teilnahme ausgeschlossen. Experiment 3 beinhaltet ein 3 x 3 Design mit dem dreistufigen Zwischensubjektfaktor *Bedingung* (Realgruppen- vs. Nominalgruppen- vs. Individualbedingung) und dem dreistufigen Innersubjektfaktor *Steuerungsphase* (Steuerungsphase 1 vs. Steuerungsphase 2 vs. Steuerungsphase 3). In einigen Analysen auf Gruppenebene kommt der dreistufige Innersubjektfaktor *Gruppenmitglied* (bestes vs. mittleres vs. schlechtestes) in Bezug auf die individuelle Leistung in der Steuerungsphase 1 hinzu.

## **6.3 Methode**

### **6.3.1 Steuerungsaufgabe und Leistungsmaß**

In Experiment 3 wurde erneut das dynamische Kühlsystemszenario im Atomkraftwerkkontext aus Experiment 1 verwendet. Ich habe mich für dieses Szenario zur Replikation der Ergebnisse und zum Nachweis von stabilem G-I Transfer entschieden, da die durchschnittliche Steuerungsdauer pro Phase geringer ist als bei dem Produktions- und Lagermanagementszenario. Wie zuvor wurde das Experiment mit der Experimentalsoftware ALEX (Schlemmer, 2009) programmiert und computergestützt durchgeführt. Das Aufgabenszenario, die genauen Instruktionen und die dem Szenario zugrunde liegenden Systemgleichungen sind detailliert unter 4.3.1 beschrieben. Als Leistungsmaß wurde erneut der MAPE-Wert in Bezug auf die zu erreichende Optimaltemperatur im Kühlsystem berechnet, der je nach Analyse auf den individualbasierten bzw. den gruppenurteilsbasierten Steuerungsergebnissen beruhte.

### **6.3.2 Experimenteller Ablauf**

Der experimentelle Ablauf war bis auf 2 Änderungen identisch mit dem experimentellen Ablauf des Experimentes 1, der unter Abschnitt 4.3.2 sowie im Anhang A erläutert wird. Die erste Änderung bezog sich auf den NEO-FFI von Borkenau und Ostendorf (1993). Da sich in Experiment 1 weder Hinweise auf eine extraversionbasierte Gewichtung der individuellen Beiträge im Realgruppenkontext noch auf einen Zusammenhang der Persönlichkeitsfaktoren mit der gruppenbasierten Steuerungsleistung fanden, wurde dieser Fragebogen als irrelevant für die zu untersuchende Fragestellung erachtet und in Experiment 3 ausgelassen. Die zweite Änderung bezog sich auf die Erweiterung des Designs um eine weitere individuelle Steuerungsphase (Steuerungsphase 3). Die mündlich durch die Versuchsleiter gegebenen bzw. schriftlich am Computer dargebotenen Instruktionen, die sich auf die Anzahl der Steuerungsphasen bezogen, wurden entsprechend angepasst. So wussten die Probanden, dass sie nach einer individuellen Systemexplorationsphase, einer ersten individuellen Steuerungsphase mit Zielvorgabe und einer zweiten Steuerungsphase im Realgruppenkontext, computervernetzten Nominalgruppenkontext oder Individualsetting noch ein weiteres Mal das System individuell steuern würden (vgl. Abbildung 15). Die Probanden in der Realbedingung wurden vor der Steuerungsphase 3 von der Versuchsleitung zurück an Einzelarbeitstische gesetzt. Nach jeder Steuerungsphase erfolgte die Erhebung des



individuellen Wohlbefindens bzw. der individuellen Motivation, das System zu verstehen und zu steuern wie in Anhang A im genauen Wortlaut beschrieben.

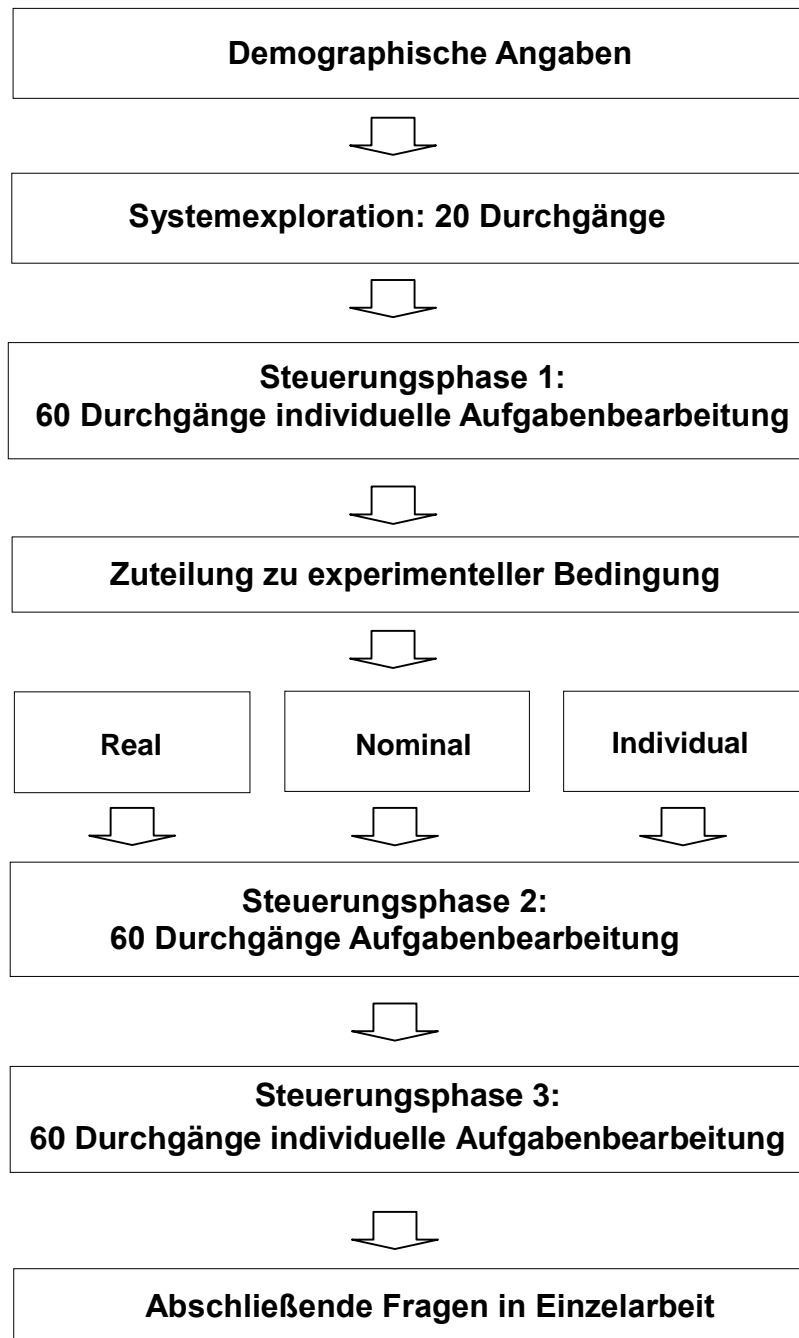


Abbildung 15. Schematische Darstellung des Ablaufes von Experiment 3.

Auch die Ankündigung bzw. Umsetzung der Grundvergütung und Bonuszahlung wurde aufgrund der zusätzlichen Steuerungsphase und damit einhergehender längerer Bearbeitungsdauer angepasst. Die Probanden erhielten 10 € Grundvergütung (statt 5 in Experiment 1) und jeweils bis zu 5 € leistungsabhängigen Bonus für die Steuerung der

Phasen 2 und 3. Maximal konnten die Probanden also 20 € für die Experimentteilnahme erhalten. Die individuelle Steuerung in Phase 3 wurde ebenfalls leistungsabhängig vergütet, um das Niveau der Leistungsbereitschaft in Phase 2 aufrecht zu erhalten und Ermüdungseffekten und Motivationsverlusten durch die lange Experimentaldauer und wiederholte Aufgabenbearbeitung entgegenzuwirken. Die MAPE-Werte zur Festlegung der Bonusbereiche entsprachen denen aus Experiment 1. Die Realgruppeninteraktion wurde wie in Experiment 2 auf Video aufgezeichnet.

### 6.3.3 Erfassung des Wissenserwerbs

Das Wissen der Probanden über die Eigenschaften und Funktionsweise des Systems wurde nach der Explorationsphase, nach der ersten, nach der zweiten sowie nach der dritten Steuerungsphase erhoben. Aufgrund der Erfahrungen aus dem ersten Experiment, in dem die Gesamtzahl genannter Annahmen im Verlaufe des Experimentes stark zurückging und in dem häufig mehrere Vermutungen unstrukturiert in ein Textfeld geschrieben wurden, was die Kodierung erschwerte, wurden zwei Erweiterungen der Instruktion vorgenommen. Erstens wurden die Probanden explizit aufgefordert, alle Gedanken zur Funktionsweise des Systems, die sie während der letzten Steuerungsphase gehabt hatten zu notieren, auch wenn sie diese Annahmen schon zuvor notiert hatten. Zweitens wurden die Probanden gebeten bei mehreren Angaben im dritten Textfeld eine strukturierende Nummerierung vorzunehmen.

## 6.4 Ergebnisse und Diskussion

Vor der Testung der aufgestellten Hypothesen wurde der Einfluss von Alter, Anzahl der Studiensemester, Studienfach und Geschlecht auf die individuelle Steuerungsleistung überprüft. Das Steuerungsleistungsmaß war wie in Experiment 1 der MAPE über die Trials 2 bis 60 einer Steuerungsphase (vgl. Kapitel 4.3.1). Das Alter der Probanden zeigte einen signifikanten Zusammenhang mit der individuellen Leistung in der Steuerungsphase 1,  $r = -.11$ ,  $p = .049$ <sup>66</sup>. Da das Durchschnittsalter in den Bedingungen vergleichbar war, wurde die Variable bei den weiteren Analysen nicht berücksichtigt,  $F(1,311) = 1.610$ ,  $p = .201$ ,  $\eta^2_p = .010$ . Die Korrelation zwischen Alter und Leistung in der Individualbedingung in Steuerungsphase 2 war nicht signifikant,  $r(104) = -.10$ ,  $p = .291$ . Auch die Korrelationen zwischen der Anzahl der bisherigen Studiensemester und der Steuerungsleistung in der Phase 1 bei Betrachtung aller Probanden war signifikant,  $r(314) = -.14$ ,  $p < .001$ . Allerdings

---

<sup>66</sup> Alle Korrelationen in diesem Abschnitt wurden zweiseitig getestet.

gab es keine Konfundierung von Studiendauer und Bedingung,  $F(2,306) = .697$ ,  $p = .499$ ,  $\eta^2_p = .01$ . In der Individualbedingung der Steuerungsphase 2 lag keine signifikante Korrelation von MAPE und Studiensemester vor,  $r(104) = -.14$ ,  $p = .153$ . Die Art des Studienfaches wurde in die fünf Kategorien Naturwissenschaften, Geisteswissenschaften, Psychologie, Wirtschaftswissenschaften und eine Restkategorie<sup>67</sup> umkodiert. Dabei bestand kein signifikanter Einfluss auf die individuelle Leistung der 315 Probanden in der Steuerungsphase 1,  $F(4,309) = 1.12$ ,  $p = .345$ ,  $\eta^2_p = .01$ . Abschließend wurden Geschlechtsunterschiede bei der individuellen Steuerung in zwei t-Tests überprüft. Bei der individuellen Steuerungsleistung in Phase 1 zeigte sich (im Gegensatz zum Experiment 1) kein signifikanter Geschlechtsunterschied,  $t(312) = 1.79$ ,  $p = .075$ ,  $d = 0.19$ . Die Steuerungsleistung der Männer ( $M = 1.04$ ,  $SD = .57$ ) war nur deskriptiv besser als die der Frauen ( $M = 1.15$ ,  $SD = .55$ ) in der Stichprobe. In der Steuerungsphase 2 fiel die Leistung der 55 weiblichen Probanden in der Individualbedingung wie in Experiment 1 ( $M = 1.08$ ,  $SD = .62$ ) signifikant schlechter aus als die Leistung der männlichen ( $M = .74$ ,  $SD = .52$ ),  $t(102) = 2.94$ ,  $p = .004$ ,  $d = 0.58$ . Für die weiteren Analysen waren Geschlechtsunterschiede unproblematisch, da keine Unterschiede der Geschlechtsverteilung in den Bedingungen bestanden,  $\chi^2(2, N = 312) = .487$ ,  $p = .787$ ,  $\phi = .40$ .

### 6.4.1 Gruppenleistung

Entsprechend der Vorgehensweise in Experiment 1, die unter 4.4.1 detailliert beschrieben ist, wurde untersucht, ob Prozessgewinne durch gruppenbasierte Systemsteuerung auftraten und interagierenden Dreiergruppen eine bessere Leistung erbrachten als die Nominalgruppen mit gemitteltem Systemverlauf in der Nominalbedingung und die Nominalgruppen mit individuellem Systemverlauf in der Individualbedingung.<sup>68</sup> Als Leistungsmaß wurde wieder die Differenz aus der Startkompetenz der Gruppen (gemessen über die pro Gruppe gemittelte individuelle Steuerungsleistung in Phase 1) und der Leistung in der zweiten Steuerungsphase berechnet. Letztere wurde in der Realbedingung mit dem gruppenurteilsbasierten MAPE-Wert über die Steuerungstrials 2 bis 60 erfasst. In der Nominalbedingung resultierten die MAPE-Werte aus dem vom Servercomputer ermittelten Gruppenmittelwert der drei individuellen Steuerungseingaben für die Trials 2 bis 60 der zweiten Steuerungsphase. In

---

<sup>67</sup> Die häufigsten der Restkategorie der Studienfächer zugeordneten Angaben waren Sportwissenschaften, Agrarwissenschaften und Geographie bzw. fehlende Angaben.

<sup>68</sup> Auf den explorativen Vergleich der gruppenbasierten Steuerung mit den Leistungsbenchmarks wurde verzichtet, da in Experiment 3 ausschließlich der Nachweis von G-I Transfer und die Replikation der hypothesenbezogenen Ergebnisse fokussiert wurden.

der Individualbedingung wurde aus den individuellen Steuerungsverläufen der Gruppenmitglieder der mittlere MAPE berechnet. Positive Werte des Differenzmaßes zwischen Startkompetenz und Leistung in der zweiten Steuerungsphase implizieren eine Leistungszunahme. Eine ANOVA mit dem Zwischensubjektfaktor *Bedingung* (Realbedingung vs. Nominalbedingung vs. Individualbedingung) und dem Differenzmaß aus der Startkompetenz der Gruppen und der gruppenbezogenen Leistung in der zweiten Steuerungsphase als abhängiger Variable zeigte einen signifikanten Effekt,  $F(2,102) = 6.53$ ,  $p = .002$ ,  $\eta^2_p = .11$ . Der Bonferroni-adjustierte Kontrast zwischen dem Leistungsmaß der Realbedingung ( $M = .47$ ,  $SD = .37$ ) und der Individualbedingung ( $M = .21$ ,  $SD = .31$ ) war signifikant,  $p = .001$ ,  $d = 0.75$ . Die Hypothese 1a, die besagt, dass die Systemsteuerungsleistung real interagierender Gruppen der Steuerungsleistung von Nominalgruppen mit individuellem Systemverlauf in der Individualbedingung überlegen ist, wird daher angenommen. Der vorhergesagte Unterschied zwischen der Realbedingung und der Nominalbedingung ( $M = .38$ ,  $SD = .17$ ) war jedoch nur deskriptiv vorhanden und nicht statistisch signifikant,  $p = .524$ ,  $d = 0.33$ . Entsprechend kann die Hypothese 1b an dieser Stelle nicht unterstützt werden.

Die deskriptiv bessere mittlere Realgruppenleistung gegenüber der Nominalgruppenleistung bei gemitteltem Systemverlauf stabilisiert sich wie in Abbildung 16 zu sehen in der zweiten Hälfte der Steuerungsphase 2. Die hohen, über die Bedingungen nahezu identischen MAPE-Werte in den ersten vier Trials sind maßgeblich durch den Ausgangszustand des Systems bedingt.

Die mittlere Steuerungsleistung in der Individualbedingung, die ab Trial 17 deskriptiv auf einem schlechteren Leistungsniveau verblieb als in der Real- und auch in der Nominalbedingung, war (knapp) nicht signifikant von der Nominalgruppenleistung verschieden,  $p = .087$ . Der Kontrast wies allerdings eine mittlere Effektstärke mit Cohen's  $d = 0.66$  auf.<sup>69</sup>

---

<sup>69</sup> Ein Vergleich der Steuerungsleistung zwischen den drei Bedingungen für die zweite Hälfte der Phase 2 mit dem MAPE der Trials 31 bis 60 entsprach vom Ergebnismuster der Analyse der Gesamtphase,  $F(2,102) = 5.52$ ,  $p = .005$ ,  $\eta^2_p = .10$ . Die Real- ( $M = .36$ ,  $SD = .53$ ) und Nominalgruppenleistung ( $M = .44$ ,  $SD = .30$ ) fielen beide signifikant bzw. deskriptiv besser aus als die Leistung in der Individualbedingung ( $M = .66$ ,  $SD = .33$ ),  $p = .005$  bzw.  $p = .064$ . Sie unterschieden sich aber erneut nicht signifikant voneinander,  $p = .711$ .

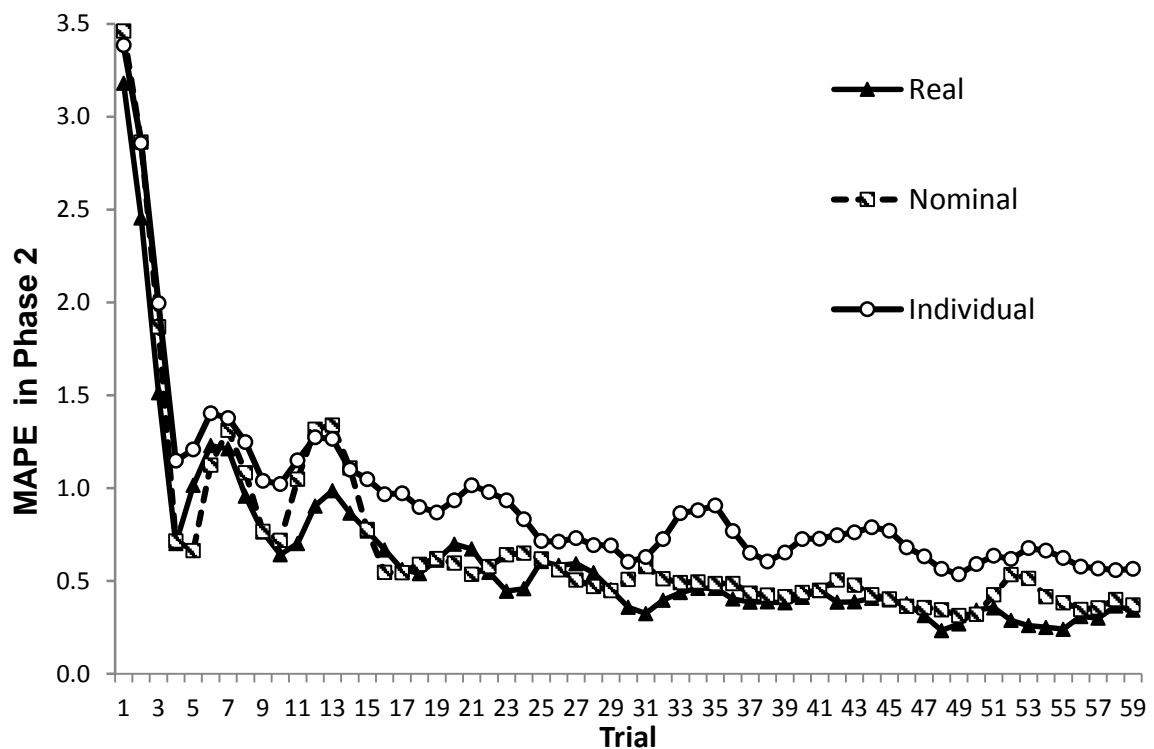


Abbildung 16. Verlauf der Temperaturkurven über die 60 Trials der Steuerungsphase 2, die aus der Steuerung in den drei Bedingungen in Experiment 3 resultierten.

#### 6.4.2 Lerneffekte auf Gruppenebene

In Experiment 1 entsprach die Gruppenleistung der Realgruppen in der zweiten Steuerungsphase dem Gruppenpotenzial aus dem Mittelwert der Individualurteile. Mit anderen Worten gelang es den Gruppen nicht durch die gemeinsame Systemsteuerung Koordinationsgewinne bzw. Synergieeffekte zu erzielen. Der Vergleich mit einer Reihe von normativen Modellen zeigte entsprechend, dass das Average Modell die beste Passung auf die Gruppenleistungsdaten hatte. Im Folgenden wird untersucht, ob dieser Befund anhand der Stichprobe in Experiment 3 replizierbar ist. Da es sich ebenfalls um eine studentische Stichprobe handelt und das identische dynamische System zu steuern ist, nehme ich an, dass die Hypothese 2a zutrifft. Diese besagt, dass die auf Gruppenurteilen basierende Steuerungsleistung der aus dem Mittelwert der individuellen Urteile der Gruppenmitglieder resultierenden Steuerungsleistung entspricht und real interagierende Gruppen keine Koordinationsgewinne realisieren.

In Experiment 3 gab es zwei methodische Änderungen gegenüber Experiment 1, die möglicherweise die Motivation der Probanden in der Realbedingung steigern könnten.

Erstens wurde die Gruppeninteraktion auf Video für spätere Analysezwecke aufgezeichnet, was allen Probanden bewusst war. Das könnte die subjektive Bedeutsamkeit des Interaktionsprozesses erhöht haben und dazu beigetragen haben, dass die Realgruppen sich stärker bemühten, die Individualurteile leistungsförderlich zu Gruppenurteilen zu kombinieren und z.B. das beste Gruppenmitglied oder akkurateste Urteil zu identifizieren. Zweitens konnten die Probanden einen doppelt so hohen maximalen Bonus durch gute Steuerung erzielen wie die Probanden in Experiment 1. Sie bekamen maximal 5 € für eine gute Steuerung in Phase 2 und weitere 5 € für eine gute individuelle Steuerung in Phase 3. Dies könnte ebenfalls die Motivation zu einer intensiven Interaktion in der Gruppe gesteigert haben, um davon noch in Phase 3 individuell profitieren zu können. Falls Realgruppen durch eine solche (meiner Meinung nach unwahrscheinliche) Motivationssteigerung systematisch besser das beste Gruppenmitglied oder das akkurateste Einzelurteil pro Trial erkennen könnten, würde dies für das Zutreffen der Hypothese 2b sprechen. Diese geht davon aus, dass die auf Gruppenurteilen basierende Steuerungsleistung die aus dem Mittelwert der individuellen Urteile der Gruppenmitglieder resultierende Steuerungsleistung übertrifft und in Gruppen Koordinationsgewinne auftreten. Zur Testung der Hypothesen wird die Gruppenleistung mit dem Average Model verglichen. Entsprechend zur Auswertung in Experiment 1 wird auch die Passung des Accuracy Models, Initial Best Member Models und des True Best Member Models auf die Gruppenleistungsdaten untersucht. Die inhaltliche und methodische Erläuterung der Modelle findet sich unter 4.4.2.

**Gruppenleistung vs. Average Model.** Ob die Realgruppen im Zeitverlauf der Steuerungsphase 2 Prozessgewinne realisierten und das Gruppenpotenzial übertrafen, wurde in einer MANOVA mit einem 2 (*Leistungsmaß*: Realgruppenleistung vs. Average Model) x 4 (*Trialblock*: 1 bis 4) Design überprüft. Der Faktor Leistungsmaß war nicht signifikant,  $F(1, 34) = .67, p = .417, \eta^2_p = .02$ . Die Realgruppenleistung-MAPE-Werte waren bis auf die gerundete dritte Nachkommastelle für alle vier Blöcke der Phase 2 identisch mit der erwarteten Leistung bei Mittelung der individuellen Urteile (Trialblock 1:  $M = 1.22, SD = .27$ , Trialblock 2:  $M = .58, SD = .64$ , Trialblock 3:  $M = .40, SD = .59$ , Trialblock 4:  $M = .31, SD = .50$ ). Der Faktor Trialblock war signifikant,  $F(2.28, 77.672) = 53.60, p < .001, \eta^2_p = .61$ . Die MAPE-Werte fielen von Block 1 zu Block 2 signifikant und zwischen den folgenden Trialblöcken deskriptiv stark ab, was der zu erwartenden Annäherung der Temperatur an den optimalen Zielwert über die Zeit entsprach. Dieser Verlauf wurde durch einen linearen sowie einen quadratischen Innersubjektcontrast beschrieben,  $F(1,34) = 120.17, p < .001, \eta^2_p = .78$  bzw.  $F(1,34) = 19.20, p < .001, \eta^2_p = .36$ . Bei gleichem Verlauf der Leistungsmaße lag erwartungsgemäß keine signifikante Interaktion zwischen Trialblock

und Leistungsmaß vor,  $F(1.30, 44.06) = 1.84$ ,  $p = .181$ ,  $\eta^2_p = .05$ . Die ANOVA repliziert somit das Ergebnis aus Experiment 1 und spricht für die Annahme der Hypothese 2a: Die Gruppeninteraktion resultierte nicht in Koordinationsgewinnen bei der Steuerung des Kühlsystemszenarios, sondern entsprach dem dynamisch im Steuerungsverlauf angepassten Gruppenpotenzial bzw. dem Average Model. Als nächstes wurde überprüft, inwiefern die normativen Standards des Accuracy Models, des Initial Best Member Models und des True Best Member Models auf die Gruppenleistungsdaten passen.

***Root Mean Squared Difference (RMSD): Passung der Vergleichsmodelle.***

Abschließend wurde anhand des RMSD überprüft, wie gut die normativen Modelle jeweils die Realgruppenleistung beschrieben. Das Average Model wies mit einem RMSD von .03 ganz deutlich die beste Passung auf, gefolgt vom Initial Best Member Model mit .54, dem True Best Member Model mit .62 und zuletzt dem Accuracy Model mit .68. In separaten Wilcoxon-Tests mit dem RMSD des Average Models und jeweils einem der anderen normativen Modelle zeigte sich, dass der RMSD in allen Fällen signifikant niedriger ausfiel, alle  $Z = 5.16$ , alle  $p < .001$ . Das Average Model hatte somit die beste Passung, um die Steuerungsleistung der Realgruppen zu beschreiben. Somit wird die Hypothese 2a angenommen, die davon ausgeht, dass keine Koordinationsgewinne auftreten und die auf Gruppenurteilen basierende Steuerungsleistung gleich oder schlechter ausfällt als die aus dem Mittelwert der individuellen Urteile der Gruppenmitglieder resultierende Steuerungsleistung.

***Vergleich der Realgruppenleistung und der Modellpassungen zwischen den Experimenten 1 und 3.*** Das Befundmuster zur Passung der durch die normativen Modelle vorhergesagten Leistung in Relation zur Gruppenleistung in der Realbedingung ist weitgehend identisch zwischen den Experimenten 1 und 3. Erstens zeigte das Average Model in beiden Untersuchungen den besten Fit und die Gruppenleistung entsprach dem Gruppenpotenzial. Zweitens fiel die Gruppenleistung in beiden Fällen signifikant schlechter als nach dem Accuracy Model aus. Bei Realisierung von G-IG Transfer oder G-G Transfer hätten die Gruppen ihre Leistung demnach noch bedeutsam steigern können. Drittens entsprach die Gruppenleistung in beiden Experimenten dem Leistungsniveau des besten Gruppenmitglieds der individuellen Steuerungsphase 1. Beim Vergleich des True Best Member Models (gemäß der Individualleistung in Phase 2) und der Gruppenleistung unterschieden sich die Ergebnisse: Während die Leistung des True Best Member in Experiment 1 der Gruppenleistung (sowie dem Average Model) signifikant überlegen war, lagen in Experiment 3 keine Unterschiede zur Gruppenleistung bzw. dem Average Model vor. Dennoch wies das Average Model auch in Experiment 3 mit einem RMSD von .03 eine

signifikant bessere Passung auf als das True Best Member Model mit einem RMSD von .62. Die deskriptiven Statistiken und RMSD-Werte für alle Modelle sind in Tabelle 10 dargestellt.

Tabelle 10: *Resultierende MAPE-Werte und RMSD unter den normativen Vergleichsmodellen in Experiment 1 und 3 im Vergleich*

Vergleichsmodell	<i>n</i>	Experiment 1		Experiment 3	
		<i>M(SD)</i>	RMSD	<i>M(SD)</i>	RMSD
Average	35	.74 (.42)	.03	.62 (.44)	.03
Accuracy	35	.65 (.37)	.80	.55 (.36)	.68
Initial Best Member	35	.73(.41)	.64	.63(.45)	.54
True Best Member	35	.71(.40)	.82	.60(.42)	.62

*Anmerkung:* Die auf zwei Nachkommastellen gerundete Realgruppenleistung gemäß des MAPE in Phase 2, auf der die RMSD-Abweichungsmaße basieren, entsprach bis auf die zweite Nachkommastelle den hier aufgeführten deskriptiven Werten des Average Models.

Daraus wird ersichtlich, dass die fehlende Überlegenheit des True Best Member Models gegenüber dem Average Model nicht auf einer systematisch schlechteren Individualleistung gegenüber Experiment 1 beruht. Stattdessen zeigt sich, dass das Leistungslevel auf Basis des Gruppenpotenzials in Experiment 3 höher lag. Entsprechend unterschied sich die Realgruppenleistung signifikant zwischen den Experimenten 1 ( $M = .74$ ,  $SD = .42$ ) und 3 ( $M = .62$ ,  $SD = .44$ ),  $t(68) = 1.18$ ,  $p = .243$ ,  $d = 0.28$ . Zu Beginn dieses Teilkapitels wurde die Videoaufzeichnung während der Gruppeninteraktion als möglicher Motivator zur Leistungssteigerung in Phase 2 genannt. Diese Erklärung wird nicht durch die Befundlage gestützt, dass bereits der MAPE der Startkompetenz für die erste Steuerungsphase in Experiment 3 mit 1.10, ( $SD = .16$ ) signifikant niedriger ausfiel als in Experiment 1 (1.30,  $SD = .28$ ),  $t(68) = 2.69$ ,  $p = .009$ ,  $d = 0.65$ . Nicht auszuschließen ist, dass die Probanden im dritten Experiment von Anfang an motivierter waren, da Ihnen ein doppelt so hoher Bonus für gute Steuerungsleistung in Aussicht gestellt worden war. Die entsprechenden Experiment übergreifenden Analysen zu der subjektiv berichteten Motivation der Gruppenmitglieder sind in Anhang N erläutert.



### 6.4.3 Lerneffekte auf individueller Ebene

Um das Auftreten von I-I Transfer (Übungseffekten) in der Individualbedingung bzw. von individueller Leistungsverbesserung in der Realbedingung und in der Nominalbedingung zu untersuchen, wurde analog zum Experiment 1 zunächst für jeden Probanden die Differenz aus dem individuellen MAPE-Wert der ersten Steuerungsphase und dem aus den individuellen Steuerungsurteilen resultierenden MAPE-Wert der zweiten Steuerungsphase berechnet. Positive Werte des Differenzmaßes implizieren eine individuelle Leistungssteigerung von der ersten zur zweiten Steuerungsphase.<sup>70</sup> Die Lerneffekte wurden auf Gruppenebene ausgewertet, indem der Faktor Gruppenmitglied als Innersubjektfaktor in die Analyse aufgenommen wurde.<sup>71</sup> Einleitend zu Experiment 3 wurde unter 6.1 erläutert, dass dieses Leistungsdifferenzmaß lediglich eindeutige Aussagen zum I-I Transfer in der Individualbedingung ermöglicht. Eine signifikant größere Leistungsverbesserung in der Realbedingung gegenüber den anderen Bedingungen darf hingegen nicht als reiner G-I Transfer interpretiert werden, da die Individualleistung in Phase 2 bei gruppenurteilsbasierten Systemzuständen erbracht wurde. Damit profitierte auch die Individualleistung vom statistischen Aggregationsvorteil der Gruppenurteile bzw. der Gruppenmittelwerte in der Nominalbedingung. Daher wird über die Analysen im ersten Experiment hinaus im Abschnitt 6.4.4 anhand eines Leistungsdifferenzmaßes zwischen der ersten individuellen und einer dritten individuellen Steuerungsphase untersucht, ob über die Gruppeninteraktion hinaus stabiler G-I Transfer in der Realbedingung auftrat.

Zunächst erfolgt jedoch in diesem Abschnitt die Ergebnisdarstellung zum Differenzmaß zwischen der ersten und zweiten Steuerungsphase, um die Replizierbarkeit der Befunde aus Experiment 1 zu überprüfen und einen direkten Vergleich der individuellen Leistungsdifferenz der Phasen 2 und 3 anhand derselben Stichprobe zu ermöglichen. Es wurde eine 3 (*Bedingung*: Realbedingung vs. Nominalbedingung vs. Individualbedingung) x 3 (*Gruppenmitglied*: bestes vs. mittleres vs. schlechtestes) ANCOVA mit der

---

<sup>70</sup> Es wurde pro Gruppenmitglied der MAPE der Phase 2 vom MAPE der Phase 1 subtrahiert. Ein in Relation zur Phase 1 kleinerer MAPE, der eine bessere Steuerungsleistung bedeutet, ging daher mit einem positiven Wert des Differenzmaßes einher.

<sup>71</sup> Wie unter 4.5 erläutert, waren die Urteile innerhalb einer Gruppe nicht unabhängig voneinander, da die individuellen Steuerungsurteile in der Realbedingung und in der Nominalbedingung alternierend zu den Gruppenurteilen bzw. den vom Serverrechner zurückgemeldeten Gruppenmittelwerten abgegeben wurden und der jeweilige Systemzustand durch die gruppenbasierten Urteile hervorgerufen wurde. Dies machte eine Analyse auf Gruppenebene erforderlich.

Startkompetenz<sup>72</sup> der Gruppen als Kovariate berechnet. Der Status der Gruppenmitglieder resultierte aus der relativen individuellen Leistung in der ersten Steuerungsphase. Die ANCOVA zeigte einen signifikanten Haupteffekt für die Bedingung,  $F(2, 101) = 6.75$ ,  $p = .002$ ,  $\eta^2_p = .12$ . Bonferroni-korrigierte post-hoc Vergleiche bescheinigten eine signifikant größere Leistungsdifferenz in der Realbedingung als in der Individualbedingung,  $p = .001$ ,  $d = 0.74$ . Die paarweisen Vergleiche der individuellen Leistungssteigerung in der Nominalbedingung mit der Real- und der Individualbedingung waren nicht signifikant,  $p = .279$ ,  $d = 0.43$  bzw.  $p = .180$ ,  $d = 0.52$ . Post-hoc Einzel-t-Tests gegen Null zeigten in allen Bedingungen eine signifikante individuelle Leistungsverbesserung, alle  $p < .001$ , alle  $d > .71$ . Die genauen Ergebnisse der t-Tests sind in Tabelle 11 zu sehen. Die deskriptiv größte individuelle Leistungssteigerung bestand in der Realbedingung ( $M = .47$ ,  $SD = .36$ ), gefolgt von der Differenz in der Nominalbedingung ( $M = .35$ ,  $SD = .17$ ). Die Übungseffekte in der Individualbedingung fielen deskriptiv am geringsten aus ( $M = .22$ ,  $SD = .31$ ). In der ANCOVA zeigte sich darüber hinaus ein signifikanter Haupteffekt für die Kovariate Startkompetenz,  $F(1, 101) = 0.75$ ,  $p = .388$ ,  $\eta^2_p = .01$ .<sup>73</sup> Der Haupteffekt für den Faktor Gruppenmitglied war nicht signifikant,  $F(1.78, 179.23) = 0.33$ ,  $p = .690$ ,  $\eta^2_p = .00$ . Der Interaktionseffekt von Gruppenmitglied x Bedingung war jedoch signifikant,  $F(3.55, 179.23) = 24.81$ ,  $p < .001$ ,  $\eta^2_p = .27$ .<sup>74</sup> Wie in Abbildung 17 und Tabelle 11 ersichtlich ist, fiel der individuelle Fertigkeitsszuwachs in der Individualbedingung für die Gruppenmitglieder mit mittlerer und schlechtester Ausgangsleistung geringer aus als in der Real- und in der Nominalbedingung.

---

<sup>72</sup> Die Startkompetenz hat Einfluss auf die Höhe des individuellen Verbesserungspotenzials der Gruppenmitglieder (je geringer die durchschnittliche individuelle Ausgangsleistung ist, desto größer ist das Ausmaß der möglichen Leistungssteigerung) und wurde zur Reduktion der Fehlervarianz als Kovariate in die Analyse einbezogen.

<sup>73</sup> Im Gegensatz zu Experiment 1 konnte der signifikante Effekt der Startkompetenz hier nicht über alle Bedingungen hinweg durch höhere individuelle Fertigkeitsszuwächse in Gruppen mit niedrigerer Startkompetenz erklärt werden,  $r(105) = .04$ ,  $p = .662$ . Für die Individualbedingung fand sich aber eine signifikante Korrelation mit  $r(35) = .42$ ,  $p = .013$ .

<sup>74</sup> Der lineare Innersubjektcontrast für Gruppenmitglied x Bedingung war mit  $F(2, 101) = 34.64$ ,  $p < .001$  und  $\eta^2_p = .33$  signifikant. Außerdem fand sich eine signifikante Interaktion von Gruppenmitglied und Startkompetenz,  $F(1.78, 179.23) = 8.05$ ,  $p = .001$ ,  $\eta^2_p = .07$ , die ebenfalls durch einen linearen Innersubjektcontrast qualifiziert wurde,  $F(1, 101) = 14.12$ ,  $p < .001$ ,  $\eta^2_p = .02$ .

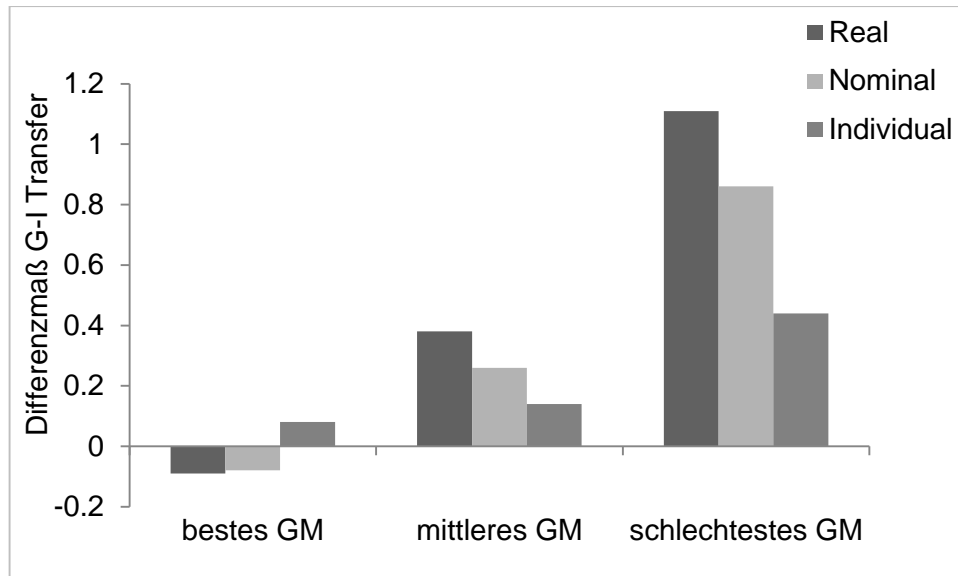


Abbildung 17. Ausmaß der individuellen Leistungssteigerung für das gemäß der Individualleistung in Phase 1 beste, mittlere und schlechteste Gruppenmitglied (GM) in Experiment 3. G-I Transfer wurde über ein Differenzmaß aus dem MAPE der Phase 1 und dem MAPE der Phase 2 bestimmt.

Der mittlere Fertigungszuwachs des besten Mitglieds in der Individualbedingung lag mit  $.08$  ( $SD = .20$ ) im positiven Bereich, wohingegen sich die besten Realgruppenmitglieder deskriptiv und die besten Nominalgruppenmitglieder signifikant verschlechterten. Die detaillierten Analyseergebnisse in Form von post-hoc Einzel- $t$ -Tests gegen Null für das in Phase 1 beste, mittlere und schlechteste Gruppenmitglied sind in Tabelle 11 dargestellt.

Das mittlere Mitglied zeigte in der Realbedingung und in der Nominalbedingung einen deutlichen Leistungszuwachs, beide  $p < .001$ , wohingegen es sich in der Individualbedingung nur deskriptiv verbesserte,  $p = .098$ . Für das schlechteste Gruppenmitglied war die individuelle Leistungssteigerung, genauso wie die pro Gruppe über alle drei Gruppenmitglieder gemittelte/*gesamte* Leistungssteigerung, in allen Bedingungen signifikant.

Tabelle 11: Differenzwerte und Ergebnisse separater Einstichproben-t-Tests gegen Null für das Differenzmaß zum individuellen Leistungszuwachs in den Steuerungsphasen 2 und 3 des Experimentes 3

Bedingung	M (SD)	t(34)	p	Cohen's d
<b>Real</b>				
bestes Mitglied	-0.09 (.36)	-1.50	.144	0.25
	<b>-0.03 (.32)</b>	<b>-0.49</b>	<b>.630</b>	<b>0.09</b>
mittleres Mitglied	.38 (.37)	5.99	< .001	1.03
	<b>.39 (.43)</b>	<b>5.36</b>	<b>&lt; .001</b>	<b>0.91</b>
schlechtestes Mitglied	1.11 (.45)	14.65	< .001	2.47
	<b>1.10 (.44)</b>	<b>14.84</b>	<b>&lt; .001</b>	<b>2.51</b>
gesamt	.47 (.36)	7.58	< .001	1.31
	<b>.49 (.26)</b>	<b>11.15</b>	<b>&lt; .001</b>	<b>1.88</b>
<b>Nominal</b>				
bestes Mitglied	-.08 (.19)	-2.60	.014	0.42
	<b>.55 (.30)</b>	<b>1.09</b>	<b>.280</b>	<b>0.18</b>
mittleres Mitglied	.26 (.20)	7.81	< .001	1.30
	<b>.17 (.34)</b>	<b>2.89</b>	<b>&lt; .001</b>	<b>0.49</b>
schlechtestes Mitglied	.86 (.27)	18.66	< .001	3.19
	<b>.38 (.57)</b>	<b>4.00</b>	<b>&lt; .001</b>	<b>0.67</b>
gesamt	.35 (.17)	12.36	< .001	2.06
	<b>.20 (.25)</b>	<b>4.72</b>	<b>&lt; .001</b>	<b>0.80</b>
<b>Individual</b>				
bestes Mitglied	.08 (.20)	2.22	.033	0.40
	<b>.08 (.26)</b>	<b>1.84</b>	<b>.070</b>	<b>0.31</b>
mittleres Mitglied	.14 (.50)	1.70	.098	0.28
	<b>.17 (.60)</b>	<b>1.72</b>	<b>.090</b>	<b>0.29</b>
schlechtestes Mitglied	.44 (.56)	4.66	.001	0.79
	<b>.64 (.75)</b>	<b>5.08</b>	<b>.001</b>	<b>0.86</b>
gesamt	.22 (.31)	15.86	<.001	0.71
	<b>.30 (.38)</b>	<b>4.60</b>	<b>&lt;.001</b>	<b>0.79</b>

*Anmerkung:* Die Werte in der jeweils oberen Ergebniszeile beschreiben den individuellen Leistungszuwachs bzgl. der Phase 2. Die fett hervorgehobenen Werte in der jeweils zweiten Ergebniszeile beschreiben den individuellen Leistungszuwachs bzgl. der Phase 3 (vgl. die Erläuterungen in Abschnitt 6.4.4).

Zusammenfassend zeigen die bisherigen, zu Experiment 1 konsistenten Ergebnisse zur individuellen Fertigkeitsverbesserung von der ersten zur zweiten Steuerungsphase mit signifikanten Übungseffekten in der Individualbedingung, sowie signifikanten individuellen Leistungssteigerungen in der Real- und der Nominalbedingung. Die Leistungssteigerung der Realgruppenmitglieder geht signifikant über die Übungseffekte in der Individualbedingung hinaus. Die Hypothese 3a wird somit nur teilweise durch die Daten gestützt: Zwar fiel die individuelle Steuerungsleistung von Realgruppenmitgliedern größer aus als die Verbesserung der Steuerungsleistung von Nominalgruppenmitgliedern mit individuellem Systemverlauf, sie unterschied sich aber nicht signifikant von der Leistungsverbesserung der Nominalgruppenmitglieder mit gemitteltem Systemverlauf.

#### 6.4.4 G-I Transfer im Realgruppenkontext

Im Folgenden wurde die zentrale Analyse gerechnet, zu deren Durchführung das dritte Experiment konzipiert worden war. Es wurde überprüft, ob - wie von der Hypothese 3c vorhersagt - stabiler, über die Gruppeninteraktion hinaus vorhandener G-I Transfer in der Realbedingung auftrat. Hierzu wurde das Leistungsdifferenzmaß zwischen der ersten individuellen und der dritten individuellen Steuerungsphase untersucht. Falls stabiler G-I Transfer vorliegt, sollte die Verbesserung der individuellen Steuerungsleistung von Realgruppenmitgliedern *nach* der Gruppeninteraktion größer sein als die Verbesserung der Steuerungsleistung von Nominalgruppenmitgliedern mit individuellem Systemverlauf und Nominalgruppenmitgliedern mit gemitteltem Systemverlauf.

Die berechnete 3 (*Bedingung*: Realbedingung vs. Nominalbedingung vs. Individualbedingung) x 3 (*Gruppenmitglied*: bestes vs. mittleres vs. schlechtestes) ANCOVA mit der Startkompetenz<sup>75</sup> der Gruppen als Kovariate ergab einen Effekt der Bedingung,  $F(2,101) = 8.43$ ,  $p < .001$ ,  $\eta^2_p = .14$ . Post-hoc Kontraste zeigten hypothesenkonform sowohl einen größeren individuellen Fertigkeitsszuwachs in der Realbedingung ( $M = .49$ ,  $SD = .26$ ) gegenüber der Nominalbedingung ( $M = .20$ ,  $SD = .25$ ),  $p < .001$ ,  $d = 1.06$ , als auch gegenüber der Individualbedingung ( $M = .30$ ,  $SD = .38$ ),  $p = .012$ ,  $d = .58$ ) Die individuelle Leistungssteigerung in der Individualbedingung von Phase 1 zu Phase 3 war nicht signifikant von der Steigerung in der Nominalbedingung verschieden,  $p = 1$ ,  $d = .31$ . Mittels Einzel-*t*-Tests gegen Null wurde überprüft, ob die Höhe

---

<sup>75</sup> Die Startkompetenz hat Einfluss auf die Höhe des individuellen Verbesserungspotenzials der Gruppenmitglieder (je geringer die durchschnittliche individuelle Ausgangsleistung ist, desto größer ist das Ausmaß der möglichen Leistungssteigerung) und wurde zur Reduktion der Fehlervarianz als Kovariate in die Analyse einbezogen.

der Leistungssteigerung in der dritten Phase bedeutsam war. In allen Bedingungen fand eine signifikante individuelle Leistungsverbesserung statt, alle  $p < .001$ , alle  $d > .79$ . Die detaillierten Ergebnisse aller  $t$ -Tests für das jeweils beste, mittlere und schlechteste Gruppenmitglied jeder Bedingung sowie für die pro Gruppe gemittelte Leistungsdifferenz finden sich fett hervorgehoben in Tabelle 17 in den jeweils unteren Ergebniszeilen. Die größte individuelle Leistungssteigerung fand in der Realbedingung statt. Diese lag mit  $.49$  ( $SD = .26$ ) ungefähr auf dem individuellen Leistungsdifferenzniveau der zweiten Steuerungsphase. Überraschenderweise kehrte sich für die Nominal- und die Individualbedingung das deskriptive Ergebnismuster im Vergleich zum Differenzmaß der Phase 2 um: Die Differenz und damit der Übungseffekt in der Individualbedingung war mit  $.30$  ( $SD = .38$ ) deskriptiv größer als in der Nominalbedingung mit  $.20$  ( $SD = .25$ ). Das deutet darauf hin, dass der vermeintliche individuelle Lernzuwachs der Nominalgruppenmitglieder in der zweiten Steuerungsphase ein methodisches Artefakt war, das durch den statistischen Aggregationsvorteil der Gruppenmittelwertbildung hervorgerufen wurde und leistungsbegünstigende Systemzustände hervorrief.

In der ANCOVA war auch der Effekt der Kovariate Startkompetenz mit  $F(1, 101) = 7.75$ ,  $p = .006$ ,  $\eta^2_p = .07$  signifikant, was darauf zurückzuführen war, dass Gruppenmitglieder in Gruppen mit niedrigerer Startkompetenz größeren Lerntransfer manifestierten,  $r(105) = .25$ ,  $p = .01$ . Außerdem war der Effekt für das Gruppenmitglied signifikant,  $F(1.71, 172.71) = 6.673$ ,  $p = .003$ ,  $\eta^2_p = .06$ . Der Interaktionseffekt von Gruppenmitglied  $\times$  Bedingung war ebenfalls signifikant,  $F(3.420, 172.71) = 8.90$ ,  $p < .001$ ,  $\eta^2_p = .15$ . Der lineare Innersubjektcontrast für Gruppenmitglied  $\times$  Bedingung war mit  $F(2, 101) = 17.55$ ,  $p < .001$  und  $\eta^2_p = .26$  signifikant. Abbildung 18 veranschaulicht, dass die Leistungsverbesserung vom besten über das mittlere zum schlechtesten Gruppenmitglied in allen Bedingungen zunahm, aber die Leistungsunterschiede zwischen den Gruppenmitgliedern in der Realbedingung deutlich höher ausfielen als in den anderen beiden Bedingungen.<sup>76</sup> Somit wird die Hypothese 3c angenommen: Es liegt stabiler, über die Gruppeninteraktion hinaus vorhandener G-I Transfer in der Realbedingung vor. Die Verbesserung der individuellen Steuerungsleistung von Realgruppenmitgliedern *nach* der Gruppeninteraktion ist größer als die Verbesserung der Steuerungsleistung von Nominalgruppenmitgliedern mit individuellem Systemverlauf und

<sup>76</sup> Desweiteren fand sich eine signifikante Interaktion von Gruppenmitglied und Startkompetenz,  $F(1.71, 172.71) = 13.88$ ,  $p < .001$ ,  $\eta^2_p = .15$ , die sowohl durch einen linearen als auch durch einen quadratischen Innersubjektcontrast qualifiziert wurde,  $F(1, 101) = 20.78$ ,  $p < .001$ ,  $\eta^2_p = .17$  bzw.  $F(1, 101) = 5.14$ ,  $p = .009$ ,  $\eta^2_p = .17$ .

Nominalgruppenmitgliedern mit gemitteltem Systemverlauf. Sowohl in der Individual- als auch in der Nominalbedingung sind signifikante Übungseffekte vorhanden. Allerdings lag für den G-I Transfers von Phase 2 zu Phase 3 in der Realbedingung ein stabiles Leistungsniveau vor. In der Nominalbedingung folgte auf die artifizielle individuelle Leistungssteigerung durch Mittelwert basierte leistungsförderliche Systemzustände hingegen ein „Einbruch“ der Individualleistung, sobald die Probanden in Phase 3 wieder auf sich allein gestellt waren.

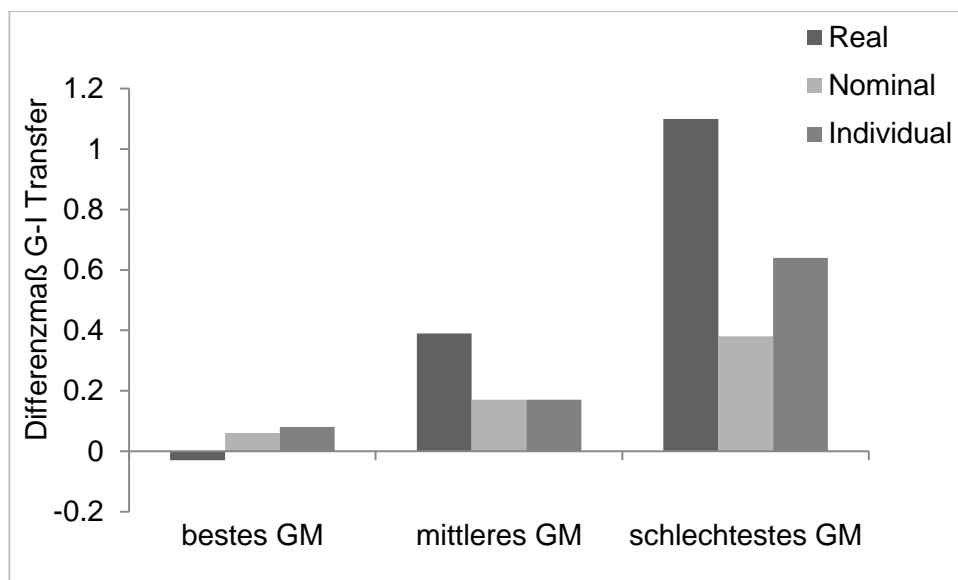


Abbildung 18. Ausmaß des G-I Transfers für das gemäß der Individualleistung in Phase 1 beste, mittlere und schlechteste Gruppenmitglied in Experiment 3. G-I Transfer wurde über ein Differenzmaß aus dem MAPE der Phase 1 und dem MAPE der Phase 3 bestimmt.

Der mittlere Fertigkeitsszuwachs des besten Mitglieds in der Individualbedingung lag mit .08 ( $SD = .20$ ) im positiven Bereich, wohingegen sich die besten Realgruppenmitglieder deskriptiv und die besten Nominalgruppenmitglieder signifikant verschlechterten. Die detaillierten Analyseergebnisse in Form von post-hoc Einzel- $t$ -Tests gegen Null für das in Phase 1 beste, mittlere und schlechteste Gruppenmitglied sind in Tabelle 17 dargestellt.

Zusammenfassend wird die Hypothese 3c angenommen: Es liegt stabiler, über die Gruppeninteraktion hinaus vorhandener G-I Transfer in der Realbedingung vor. Die Verbesserung der individuellen Steuerungsleistung von Realgruppenmitgliedern *nach* der Gruppeninteraktion ist größer als die Verbesserung der Steuerungsleistung von Nominalgruppenmitgliedern mit individuellem Systemverlauf und Nominalgruppenmitgliedern mit gemitteltem Systemverlauf.

Schließlich wurde mittels einer Mediationsanalyse explorativ untersucht, ob die im Vergleich zur Individualbedingung signifikant und gegenüber der Nominalbedingung deskriptiv höhere Leistung in der Realbedingung durch fertigungsbezogenen G-I Transfer erklärt werden konnte. Hierzu wurde als Maß für G-I Transfer die Differenz der individuell erzielten MAPE-Werte aus Phase 3 und Phase 1 herangezogen. Für die explorativen Berechnungen wurden die drei Bedingungen in die beiden Kategorien Realgruppenbedingung vs. keine Realgruppenbedingung dummy-kodiert. Die zweite Kategorie umfasste die Nominalgruppen mit gemitteltem sowie mit individuellem Systemverlauf. Im Folgenden wurden die drei nötigen Voraussetzungen für das Vorliegen einer Mediation nach Baron und Kenny (1986) überprüft. Die unabhängige Variable der Gruppenbedingung hatte erstens einen signifikanten Effekt auf die Gruppenleistung in der Steuerungsphase 2,  $\beta = -.26$ ,  $t(104) = -2.74$ ,  $p = .007$ , sowie zweitens einen signifikanten Effekt auf den fertigungsbezogenen G-I Transfer,  $\beta = .35$ ,  $t(104) = 3.73$ ,  $p < .001$ . Drittens war der fertigungsbezogene G-I Transfer bei gleichzeitigem Einfügen der Gruppenbedingung und des G-I Transfers weiterhin ein signifikanter Prädiktor für die Gruppenleistung,  $\beta = -.45$ ,  $t(104) = -4.88$ ,  $p < .001$ . Der Zusammenhang zwischen der Gruppenbedingung und der Gruppenleistung war hingegen nicht mehr signifikant, mit  $\beta = -.11$ ,  $t(104) = -1.16$ ,  $p = .249$ . Der durchgeführte Sobel Test zeigte entsprechend eine statistisch signifikante Reduktion des Regressionsgewichtes auf (vgl. Sobel, 1982),  $z = -2.96$ ,  $p = .002$  bei einseitiger Testung und  $p = .003$  bei zweiseitiger Testung. Bei alternativer Berechnung der Mediationsanalyse unter Herauspartialisierung des Effektes der Startkompetenz für das Gruppenleistungs- und das G-I Transfer-Maß blieb das Ergebnismuster der Mediation erhalten (und fiel sogar noch deutlicher aus).<sup>77</sup>

---

<sup>77</sup> Um die Leistungsmaße bezüglich der Unterschiede im durchschnittlichen individuellen Ausgangsniveau der Gruppenmitglieder zu korrigieren wurden vor der Durchführung der Mediationsanalyse die Residuen separater linearer Regressionen der Gruppenleistung sowie des G-I Transfers auf die Startkompetenz berechnet. Die Gruppenbedingung hatte einen signifikanten Einfluss auf die standardisierten Residuen der Gruppenleistung in der Steuerungsphase 2,  $\beta = -.27$ ,  $t(104) = -2.83$ ,  $p = .006$ , sowie einen signifikanten Effekt auf die standardisierten Residuen des wissensbezogenen G-I Transfers,  $\beta = .37$ ,  $t(104) = 4.01$ ,  $p < .001$ . Bei gleichzeitigem Einfügen der Gruppenbedingung und des G-I Transfers, war der fertigungsbezogene G-I Transfer weiterhin ein signifikanter Prädiktor für die Gruppenleistung,  $\beta = -.67$ ,  $t(104) = -8.45$ ,  $p < .001$ . Der Zusammenhang zwischen der Gruppenbedingung und der um das Startkompetenzniveau korrigierten Gruppenleistung war hingegen nicht mehr signifikant, mit  $\beta = -.02$ ,  $t(104) = -0.30$ ,  $p = .762$ . Ein Sobel Test wies eine statistisch signifikante Reduktion des Regressionsgewichtes nach mit  $z = -3.62$  und  $p < .001$  bei zweiseitiger sowie bei einseitiger Testung nach.



### 6.4.5 Wissenserwerb bei der Systemsteuerung

Das Wissen der Probanden über die Eigenschaften und Funktionsweise des Kühlsystemszenarios wurde zu vier Messzeitpunkten erhoben: Nach der Systemexplorationsphase, nach der ersten, individuellen Steuerungsphase, nach der zweiten Steuerungsphase im Kontext einer der Bedingungen sowie nach der dritten, individuellen Steuerungsphase. Die Probanden wurden aufgefordert alle ihre konkreten Annahmen über den Zusammenhang von Stellradeinstellung und Kühltemperatur während der vorangehenden Steuerungsphase im offenen Antwortformat zu notieren (vgl. 4.3.3 für genaue Erläuterungen der Instruktionen und Vorgehensweise bei der Auswertung). Wie für die Experimente 1 und 2 wurde pro Steuerungsphase und Proband jeweils die Gesamtsumme an Annahmen, die Summe richtiger Annahmen sowie der Prozentsatz richtiger Annahmen berechnet. Die Interraterreliabilität der zwei Kodierer für die Summenscores bezüglich der Kodierung *richtig* vs. *falsch* betrug  $Kappa = .76$  ( $SE = .01$ ),  $p < .001$ . Die Interraterreliabilität für die inhaltliche Kodierung bei anschließender Rekodierung in die Kategorien richtig vs. falsch ergab  $Kappa = .65$  ( $SE = .01$ ),  $p < .001$ . Beide Kennwerte sind nach Landis und Koch (1977) als eine substantielle Übereinstimmung (*substantial agreement*) zu bewerten. Genauso wie für die Experimente 1 und 2 wurde pro Proband jeweils ein Differenzmaß aus der ersten und zweiten Steuerungsphase für jede der drei abhängigen Variablen berechnet. Anhand des Differenzmaßes wurde getestet, ob sich die Quantität bzw. die Qualität des Wissens durch die Gruppeninteraktion im Vergleich zur Nominal- und Individualbedingung verändert hatte. In Tabelle 12 sind die Mittelwerte und Standardabweichungen der drei Differenzmaße zu finden.

Die ANOVA mit dem Zwischensubjektfaktor Bedingung (Realbedingung vs. Nominalbedingung vs. Individualbedingung) und der abhängigen Variablen Differenzmaß der Gesamtzahl an Annahmen war nicht signifikant,  $F(2,102) = 0.98$ ,  $p = .378$ ,  $\eta^2_p = .02$ . Trotz der erweiterten Instruktion der Probanden, gegebenenfalls schon zuvor genannte, noch gültige Annahmen erneut aufzuschreiben, reduzierte sich wie in Experiment 1 in allen Bedingungen die Gesamtzahl genannter Annahmen nach der zweiten Steuerungsphase im Vergleich zur Anzahl der Nennungen nach der ersten Steuerungsphase. Die Hypothese 4a, die davon ausgeht, dass Personen in Realgruppen mehr Annahmen über die Systemeigenschaften treffen als Nominalgruppen mit gemitteltem oder individuellem Systemverlauf wird daher nicht angenommen.

Tabelle 12: *Nach Bedingungen getrennte Mittelwerte und Standardabweichungen für die absoluten Werte der abhängigen Variablen zum Wissenserwerb in Steuerungsphase 2 sowie für das Differenzmaß aus Phase 1 und Phase 2 in Experiment 3*

Variable	<i>n</i>	Werte Phase 2 <i>M (SD)</i>	<i>n</i>	Differenzmaß <i>M (SD)</i>
Gesamtzahl Annahmen				
Realbedingung	35	1.80 (.90)	35	-.05 (1.37)
Nominalbedingung	35	1.55 (.73)	35	-.37 (1.11)
Individualbedingung	35	1.79 (.72)	35	-.41 (1.03)
Anzahl richtig				
Realbedingung	35	1.17 (.78)	35	.26 (1.19)
Nominalbedingung	35	.91 (.57)	35	-.15 (1.09)
Individualbedingung	35	1.05 (.53)	35	-.37 (.80)
Prozentsatz richtig				
Realbedingung	35	65.85 (31.87)	32	18.96 (42.77)
Nominalbedingung	35	54.92 (23.80)	30	4.64 (30.52)
Individualbedingung	34	59.31 (20.09)	29	3.36 (24.89)

*Anmerkung:* Die unvollständigen Zellbesetzungen bei der Variablen *Prozentsatz richtig* resultierten aus Fällen, wo keine Annahmen nach der Steuerungsphase 2 getroffen wurden.

Eine zweite ANOVA mit dem Zwischensubjektfaktor *Bedingung* und der abhängigen Variablen *Differenzmaß der Anzahl richtiger Annahmen* war (wie in Experiment 1 auch) signifikant,  $F(2,102) = 3.25$ ,  $p = .043$ ,  $\eta^2_p = .06$ . Die Summe richtiger Annahmen nahm in der Realbedingung deskriptiv zu und in der Nominalbedingung verblieb sie auf vergleichbarem Niveau, wohingegen sie in der Individualbedingung signifikant abnahm<sup>78</sup> (vgl. auch die

<sup>78</sup> Separate Ein-Stichproben-*t*-Tests gegen 0 zeigten für das Differenzmaß der Anzahl richtiger Annahmen in der Realbedingung einen deskriptiven, nicht von 0 verschiedenen Anstieg,  $t(34) = 0.43$  und  $p = .673$ , für die Nominalbedingung keine signifikant von 0 verschiedene Veränderung,  $t(34) =$

deskriptiven Statistiken in Tabelle 12). Bonferroni-korrigierte Kontraste zeigten einen signifikanten Unterschied zwischen der Realbedingung und der Individualbedingung,  $p = .041$ . Die Einzelvergleiche zwischen der Real- und Nominalbedingung ( $p = .318$ ) sowie zwischen der Nominal- und Individualbedingung ( $p = .813$ ) unterschieden sich nicht signifikant. Aufgrund des nicht signifikanten Unterschiedes zwischen der Real- und der Nominalbedingung wird die Hypothese 4b nicht angenommen. Zwar fällt die *Anzahl korrekter Annahmen* über Systemeigenschaften bei Realgruppenmitgliedern hypothesenkonform höher aus als bei Mitgliedern von Nominalgruppen mit individuellem Systemverlauf, sie ist den Nominalgruppen mit gemitteltem Systemverlauf jedoch nur deskriptiv überlegen.

Als drittes wurde die Hypothese 4c überprüft, die annimmt, dass der Prozentsatz korrekter Annahmen über Systemeigenschaften bei Realgruppenmitgliedern höher ausfällt als bei Mitgliedern von Nominalgruppen mit gemitteltem oder individuellem Systemverlauf. Die zugehörige ANOVA mit dem Zwischensubjektfaktor Bedingung und der abhängigen Variablen Differenzmaß des Prozentsatzes richtiger Annahmen zeigte keinen signifikanten Effekt der Bedingung,  $F(2,88) = 2.03$ ,  $p = .138$ ,  $\eta^2_p = .04$ . Das Ergebnismuster der Mittelwerte war allerdings hypothesenkonform, aber wohl aufgrund der großen Binnenstreuungen nicht signifikant. Die prozentuale Zunahme richtiger Systemannahmen nach der zweiten Steuerungsphase in der Realbedingung lag im Mittel mit 18.96 % ( $SD = 42.77$ ) über der Nominalbedingung mit 4.64 % ( $SD = 30.52$ ) und der Individualbedingung mit 3.36 % ( $SD = 24.89$ ). Somit muss die Hypothese 4c für das dritte Experiment abgelehnt werden.

In weiteren explorativen Analysen wurde überprüft, ob das Ergebnismuster der Wissensveränderung nach Phase 2 auch nach der dritten individuellen Steuerungsphase erhalten blieb. Analog zu den Berechnungen der Hypothesentestungen 4a bis 4c wurden drei Differenzmaße zum Systemwissen zwischen den Steuerungsphasen 3 und 1 bestimmt. Die deskriptiven Statistiken sind in Tabelle 13 aufgeführt.

---

0.88 und  $p = .383$  und für die Individualbedingung mit  $t(34) = 2.74$  und  $p = .010$  eine signifikante Abnahme der Anzahl richtiger Annahmen.

Tabelle 13: *Nach Bedingungen getrennte Mittelwerte und Standardabweichungen für die absoluten Werte der abhängigen Variablen zum Wissenserwerb in Steuerungsphase 2 sowie für das Differenzmaß aus Steuerungsphase 3 und Steuerungsphase 1 in Experiment 3*

Variable	<i>n</i>	Werte Phase 3 <i>M</i> ( <i>SD</i> )	<i>n</i>	Differenzmaß <i>M</i> ( <i>SD</i> )
Gesamtzahl Annahmen				
Realbedingung	35	1.43 (.79)	35	-.41 (1.20)
Nominalbedingung	35	1.39 (.73)	35	-.53 (.90)
Individualbedingung	35	1.65 (.69)	35	-.55 (1.20)
Anzahl richtig				
Realbedingung	35	1.03 (.63)	35	.12 (1.06)
Nominalbedingung	35	.79 (.57)	35	-.27 (.89)
Individualbedingung	35	.97 (.51)	35	-.45 (.86)
Prozentsatz richtig				
Realbedingung	35	71.53 (31.26)	32	17.13 (45.20)
Nominalbedingung	35	54.67 (27.87)	30	-2.47 (37.87)
Individualbedingung	34	53.24 (19.17)	29	7.85 (24.27)

Das Ergebnismuster entsprach den Berechnungen mit den Differenzmaßen aus den Phasen 2 und 1 (vgl. Tabelle 13): Erstens gab es keine Unterschiede zwischen den Bedingungen in der quantitativen Zunahme des Wissens von der ersten zur dritten Steuerungsphase (sondern deskriptiv eine geringe Abnahme),  $F(2,102) = 0.16$ ,  $p = .856$ ,  $\eta^2_p = .00$ . Zweitens unterschieden sich die Bedingungen in der Steigerung der Anzahl korrekter Wissensinhalte zwischen den Phasen 1 und 3,  $F(2,102) = 3.31$ ,  $p = .041$ ,  $\eta^2_p = .06$ , was auf einen signifikant größere Zunahme in der Realbedingung als in der Individualbedingung zurückzuführen war,  $p = .041$ .<sup>79</sup> Drittens gab es keinen Unterschied in

<sup>79</sup> Die Bonferroni-korrigierten Kontraste zwischen der Real- und Nominalbedingung ( $p = .259$ ) sowie zwischen der Nominal- und Individualbedingung ( $p = 1$ ) unterschieden sich nicht signifikant.

der Zunahme des Prozentsatzes korrekter an der Gesamtzahl geäußerter Annahmen zum Kühlsystem,  $F(2,79) = 1.89$ ,  $p = .158$   $\eta^2_p = .05$ .

Mit einer Mediationsanalyse wurde analog zum Vorgehen in Experiment 1 überprüft, ob die (gegenüber der Individualbedingung signifikant und gegenüber der Nominalbedingung deskriptiv) bessere Leistung in der Realbedingung durch G-I Transfer in Form eines korrekteren Systemwissens erklärt werden konnte. Hierzu wurde als Maß für den wissensbezogenen G-I Transfer die Differenz der Anzahl richtiger Systemannahmen aus Phase 2 und Phase 1 berechnet. Positive Werte des Differenzmaßes bedeuten eine Steigerung der Wissensqualität in Phase 2. Für die Berechnungen wurden die Bedingungen abermals in die zwei Kategorien Realgruppenbedingung vs. keine Realgruppenbedingung dummy-kodiert. Die zweite Kategorie beinhaltete somit die Nominalgruppen mit gemitteltem und mit individuellem Systemverlauf. Die unabhängige Variable der Gruppenbedingung hatte einen signifikanten Effekt auf die Gruppenleistung,  $\beta = -.26$ ,  $t(104) = -2.74$ ,  $p = .007$ , sowie einen signifikanten Effekt auf den wissensbezogenen G-I Transfer,  $\beta = .23$ ,  $t(104) = 2.39$ ,  $p = .018$ . Bei gleichzeitigem Einfügen der Gruppenbedingung und des G-I Transfers, war G-I Transfer weiterhin ein signifikanter Prädiktor für die Gruppenleistung,  $\beta = -.26$ ,  $t(104) = -2.71$ ,  $p = .008$ . Der Zusammenhang zwischen der Bedingung und der Gruppenleistung war weiterhin statistisch signifikant, wies jedoch ein reduziertes  $\beta = -.20$  auf,  $t(104) = -2.13$ ,  $p = .036$ . Die Reduktion des Regressionsgewichtes für die Variable Gruppenbedingung war statistisch signifikant, wie der durchgeführte Sobel Test zeigte (vgl. Sobel, 1982),  $z = -1.86$ .  $p = .036$  bei einseitiger Testung (bzw.  $p = .073$  bei zweiseitiger Testung). Aus diesen explorativen Berechnungen lässt sich ableiten, dass der positive Effekt des Realgruppenkontextes auf die Gruppenleistung wahrscheinlich partiell durch den wissensbezogenen G-I Transfer mediiert wurde (vgl. zur Kritik an der Interpretierbarkeit des Ergebnisses die Erläuterungen unter 4.4.4).

Abschließend wurde analog zu den zwei vorangehenden Experimenten exploriert, ob es einen Zusammenhang zwischen der Leistung in der zweiten Steuerungsphase und den Indikatoren für Systemwissen gab. Die Quantität des Wissens wies keinen Zusammenhang mit der Steuerungsleistung auf. Die Korrelation zwischen dem MAPE und der Gesamtsumme der Annahmen in Phase 2 betrug  $r = -.10$ ,  $p = .293$ . Die Anzahl richtiger Annahmen und der MAPE korrelierten hingegen signifikant mit  $r = -.43$ ,  $p < .001$ . Auch der Prozentsatz richtiger Annahmen ergab mit  $r = .57$  und  $p = .001$  einen signifikanten Zusammenhang mit der Steuerungsleistung. Eine höhere Wissensqualität ging somit mit höherer Steuerungsleistung einher.

### 6.4.6 Zusammenfassung

Experiment 3 wurde durchgeführt, um zu überprüfen, ob durch die Gruppeninteraktion bei der dynamischen Systemsteuerung G-I Transfer auftritt, der zeitstabil und über die Gruppenarbeitsphase hinaus nachweisbar ist. Die Probanden in Experiment 3 steuerten das Kühlsystem hierfür dreimal: In Phase 1 individuell, in Phase 2 in der Real-, Nominal- oder Individualbedingung und in Phase 3 noch einmal individuell. Die abhängige Leistungsvariable war wie in Experiment 1 der MAPE-Wert, der zur Überprüfung von G-I Transfer in Form eines Differenzmaßes der (auf Gruppenebene gemittelten) individuellen Steuerungsleistung zwischen den Phasen 1 und 3 betrachtet wurde. Der individuelle Fertigkeitsszuwachs in der Realbedingung war hypothesenkonform gegenüber der Nominalbedingung sowie gegenüber der Individualbedingung überlegen. Die individuelle Leistungssteigerung der Probanden in der Nominalbedingung unterschied sich nicht von den Übungseffekten in der Individualbedingung und fiel deskriptiv sogar geringer aus. Die zentrale Erkenntnis des dritten Experiments ist somit, dass die Realgruppenmitglieder einen zeitstabilen individuellen Fertigkeitssgewinn (G-I Transfer) durch die Gruppeninteraktion erzielten und auch über die Gruppensteuerungsphase hinaus von der Gruppenerfahrung für ihre Individualleistung profitierten. Die entsprechende Hypothese 3c wurde angenommen.

Die Analyse der individuellen Verbesserung der Steuerungsleistung zwischen den Phasen 1 und 2 (die nicht eindeutig als G-I Transfer interpretiert werden kann) zeigte konsistent zu Experiment 1 signifikante Übungseffekte in der Individualbedingung und Leistungssteigerungen in der Real- und der Nominalbedingung. Die Verbesserung der Realgruppenmitglieder übertraf zwar die Übungseffekte, sie unterschied sich aber nicht signifikant von der Leistungsverbesserung der Nominalgruppenmitglieder bei gemitteltem Systemverlauf. Die Hypothese 3a wurde somit nur teilweise durch die Daten gestützt. Die vorübergehende vermeintliche Steigerung der Individualleistung in der Nominalbedingung kann post hoc als Resultat der günstigen, mittelwertbasierten Systemzustände und/oder der Orientierung an den Gruppenmittelwerten als Referenzrahmen interpretiert werden. Diese temporären Vorteile, führten aber nicht zu einer Steigerung der individuellen Steuerungskompetenz, die unabhängig von der Nominalgruppensituation Bestand hatte.

Darüber hinaus wurde die Replizierbarkeit der Ergebnisse in Experiment 1 zu Prozessgewinnen, Koordinationsgewinnen und Wissenserwerb im dynamischen Atomkraftwerkszenario untersucht. Die Untersuchung von Prozessgewinnen entsprach dem Befundmuster aus Experiment 1 (und auch dem Befundmuster mit dem

Wirtschaftsszenario in Experiment 2). Die Realgruppenleistung war signifikant besser als die Steuerung der Probanden in der Individualbedingung, aber nur deskriptiv der Nominalgruppenleistung überlegen. Das Ergebnismuster zu Lerneffekten auf Gruppenebene war größtenteils übereinstimmend mit dem des ersten Experiments. Die Realgruppenleistung entsprach dem Gruppenpotenzial und es lagen keine Koordinationsgewinne vor. Das Average Model hatte wie in Experiment 1 die beste Passung auf die Realgruppendaten.

Die Auswertung zum Systemwissen der Probanden mittels Differenzmaßen aus den nach Phase 1 und Phase 2 geäußerten Systemannahmen zeigte konsistent zu Experiment 1 keinen Unterschied in der Veränderung der Quantität des Systemwissens zwischen der Real-, Nominal- und Individualbedingung. Die Summe richtiger Annahmen nahm in der Realbedingung hypothesenkonform nach der zweiten Phase zu und unterschied sich signifikant von der Wissensveränderung der Individualbedingung und deskriptiv von der Veränderung in der Nominalbedingung. In beiden Vergleichsbedingungen reduzierte sich die Anzahl richtiger Annahmen in Phase 2. Es gab keinen signifikanten Bedingungsunterschied in der Zunahme des Prozentsatzes korrekter Annahmen. Deskriptiv steigerten die Realgruppenmitglieder die Korrektheit ihres Wissensrepertoires um knapp 19% und kamen auf knapp 66% richtige Wissensinhalte nach der zweiten Steuerungsphase. Die Nominalgruppenmitglieder und Probanden in der Individualbedingung erreichten nur Verbesserungen unter 5% und hatten nach der zweiten Steuerungsphase zu knapp 55% bzw. 60% korrekte Wissensrepräsentationen des Kühlsystems. Die deskriptiv relativ große Steigerung der Wissensqualität in der Realbedingung könnte man als wissensbezogenen G-I Transfer interpretieren.<sup>80</sup>

Schließlich wurde der Befund aus Experiment 1 repliziert, dass sowohl ein Zusammenhang zwischen der Anzahl als auch zwischen dem Prozentsatz richtiger Systemannahmen und dem Steuerungserfolg bestand. Daraus folgt, dass eine leistungssteigernde Intervention darin bestehen könnte, die Probanden mit relevantem Systemwissen wie Informationen über die zeitverzögerte Systemreaktion oder Nebeneffekte von Systemeingriffen zu informieren. Die Untersuchung von Diehl und Serman (1995) zeigt allerdings, dass individuelle Steuerer trotz genauer Angaben zur Zeitverzögerung und der Gewichtung der Systemparameter im Systemgleichungssystem zu Fehlinterpretationen der Systemrückmeldungen und des Systems insgesamt tendierten.

---

<sup>80</sup> Hierbei ist zu berücksichtigen, dass die Streuung innerhalb der Bedingungen sehr groß war und der Unterschied zwischen den Bedingungen trotz der Mittelwertunterschiede nicht signifikant war.

Es bleibt zu untersuchen, ob eine größere Systemtransparenz im vergleichsweise wenig komplexen Kühlsystem eine stärkere leistungsfördernde Wirkung hätte.

### 6.5 Metaanalyse zu Prozessgewinnen

Die bisherige Ergebnisauswertung der Steuerungsleistungsdaten erfolgte separat für jedes der drei durchgeführten Experimente. Dabei zeigte sich jeweils hypothesenkonform eine deskriptive Überlegenheit der Realgruppen gegenüber den Nominalgruppen; Prozessgewinne aufgrund der Gruppeninteraktion konnten aber nicht statistisch abgesichert werden: Die Realgruppenleistung war der Leistung von Nominalgruppen mit statistischem Aggregationsvorteil durch gemittelten Systemverlauf lediglich deskriptiv, aber nicht statistisch signifikant überlegen. Mit der Stichprobengröße von 35 Gruppen pro Bedingung ging eine unzureichende Teststärke einher, um kleine bis mittlere Effekte<sup>81</sup> nachzuweisen. Um diesbezüglich eine verlässlichere Effektschätzung vornehmen zu können, habe ich eine Meta-Analyse mit festen Effekten (vgl. Rosenthal, 1991) über die drei Experimente berechnet. Hierzu wurde die Software Comprehensive Meta Analysis Version 2.0 verwendet. Als Maß für Prozessgewinne ging aus jedem Experiment die Differenz zwischen der Systemsteuerungsleistung in Phase 2 und der Startkompetenz der Gruppen aus Phase 1 ein. Die metaanalytische Berechnung über die drei Experimente zeigte für den Vergleich der Realbedingung mit der Individualbedingung erwartungsgemäß einen signifikanten Effekt von  $z = 7.10$ ,  $p < .001$ . Dieser Effekt beschreibt einen deutlichen Leistungsvorteil der interaktiven Systemsteuerung in der Gruppe gegenüber der individuellen Systemsteuerung. Die Hypothese 1a, die besagt, dass Gruppeninteraktion zu einer besseren Leistung bei der Steuerung eines dynamischen Systems als die Steuerungsleistung von Nominalgruppen mit individuellem Systemverlauf führt, kann somit (wie auch schon auf Ebene der Einzelexperimente) angenommen werden. Für den im Fokus stehenden Vergleich des Leistungsdifferenzmaßes zwischen der Realbedingung und der Nominalbedingung resultierte ebenfalls ein signifikanter Effekt mit  $z = 2.25$  und  $p = .025$ . Somit kann bei entsprechend großer zugrundeliegender Stichprobe auch ein Leistungsvorteil durch *Gruppeninteraktion* gezeigt werden, der über den wisdom of the crowds effect einer statistischen Aggregation von Einzelurteilen hinausgeht. Gemäß der Hypothese 1b führt Gruppeninteraktion bei der dynamischen Systemsteuerung somit

---

<sup>81</sup> Die Effektstärke für den Prozessgewinn-Vergleich zwischen der Realbedingung und der Nominalbedingung betrug für das Produktions- und Lagermanagementszenario in Experiment 2 Cohen's  $d = 0.53$ , und für das Atomkraftwerkszenario mit der Kühlsystemsteuerung  $d = 0.08$  (in Experiment 1) bzw.  $d = 0.33$  (in Experiment 3).



ebenfalls zu einer besseren Leistung als die Steuerungsleistung von Nominalgruppen mit gemitteltem Systemverlauf. Es treten interaktionsbedingte Prozessgewinne auf.

## 7. Allgemeine Diskussion

Die zielführende Steuerung von dynamischen Systemen wie beispielsweise Kühlreaktoren in Atomkraftwerken oder nachfrageorientierten Produktions- und Lagermanagementsystemen in Wirtschaftsunternehmen ist hochgradig relevant, um die öffentliche Sicherheit und ökonomischen Erfolg zu gewährleisten. Die bisherige Forschung zur Steuerungsleistung von komplexen, dynamischen Systemen zeigt systematische Schwächen bei individueller Steuerung auf. In der organisationalen Praxis werden häufig Gruppen eingesetzt, um Steuerungsurteile zu treffen, da meist implizit angenommen wird, dass durch Gruppeninteraktion leistungsförderliche Synergieeffekte (Larson, 2010) bzw. Prozessgewinne auftreten. Bisherige Forschungsarbeiten, die die Steuerungsleistung von Gruppen und Individuen verglichen haben, erlaubten aufgrund von Variablenkonfundierungen oder methodischer Mängel wie unvollständiger Designs keine Aussage darüber, ob Gruppen Prozessgewinne realisieren und worauf der gegebenenfalls auftretende Leistungsvorteil beruht.

Das Ziel dieser Dissertation war es in einer Serie von drei Experimenten mit zwei unterschiedlich komplexen Simulationsszenarien den ersten methodisch sauberen Nachweis von Prozessgewinnen bei der dynamischen Systemsteuerung von Gruppen zu erbringen und zugrundeliegende Lernprozesse zu identifizieren. Zu diesem Zweck wurde als methodische Neuerung eine Nominalgruppenbedingung mit gemitteltem Systemverlauf eingeführt, die den interaktionsunspezifischen Effekt des statistischen Aggregationsvorteils mehrerer Urteile für den dynamischen Systemverlauf berücksichtigte. Der Vergleich der Steuerungsleistung real interagierender Gruppen mit dieser kontinuierlich im Steuerungsverlauf bestimmten Nominalgruppenleistung stellte einen angemesseneren und konservativeren Test für das Vorliegen von Prozessgewinnen dar als der zusätzlich vorgenommene (und bisher in der Forschung bei anderen statischen Aufgabentypen übliche) Vergleich mit Nominalgruppen mit individuellem Systemverlauf und post hoc berechneter Individualleistung. Die Ergebnisse zeigen, dass die Steuerungsleistung von interagierenden Realgruppen der durchschnittlichen individuellen Steuerungsleistung (in Nominalgruppen mit individuellem Systemverlauf) überlegen war. Eine Meta-Analyse über die drei durchgeführten Experimente liefert den ersten, eindeutigen Nachweis interaktionsbedingter Prozessgewinne. Sie zeigt, dass die Realgruppenleistung nicht nur die durchschnittliche Individualleistung, sondern auch den statistischen Aggregationsvorteil von computervernetzten Nominalgruppen mit gemitteltem Systemverlauf übertraf. Sowohl in einem dynamischen Kühlsystem geringer Komplexität (Experimente 1 und 3) als auch in einem Wirtschaftssystem mit größerer Komplexität (Experiment 2) lagen keine

Koordinationsgewinne vor, die über das Gruppenpotenzial des Mittelwertes der Individualurteile der Gruppenmitglieder hinausgingen. In Experiment 3 konnte erstmals ein eindeutiger Nachweis von Group-to-Individual Transfer durch gruppenbasierte Systemsteuerung nachgewiesen werden, der über die Gruppenphase hinaus stabil war. Das Ausmaß des G-I Transfers ging über die Übungseffekte bei individueller Steuerung hinaus. Der individuelle Fertigkeitengewinn stellt in den durchgeführten Experimenten einen plausiblen Mediator für den Leistungsvorteil interagierender Gruppen bei der Steuerung dynamischer Systeme dar.<sup>82</sup> Eine explorativ durchgeführte Mediationsanalyse stützte die Annahme, dass G-I Transfer den Einfluss der Gruppenbedingung auf die Steuerungsleistung vermittelte. Aufgrund methodischer Gegebenheiten konnte jedoch kein gesicherter Nachweis der Mediation erbracht werden.<sup>83</sup>

### **7.1 Gruppenleistung und Gruppenkoordination bei der Steuerung dynamischer Systeme**

Die durchgeführte Metaanalyse über die Leistungsdaten der drei Experimente dieser Arbeit zeigte, dass Gruppen durch die Interaktion Prozessgewinne erzielen, die über die durchschnittliche individuelle Steuerungsleistung in Nominalgruppen sowie über den statistischen Aggregationsvorteil durch Mittelung mehrerer Steuerungsurteile in

---

<sup>82</sup> Da kein Nachweis für Koordinationsgewinne gefunden wurde und das Average Model, das eine Mittelungsstrategie der Individualurteile zur Festlegung der Gruppenurteile simuliert, in Relation zu den anderen normativen Modellen die beste Passung auf die Gruppenleistung aufwies, muss der Gruppenleistungsvorteil gegenüber den Vergleichsbedingungen auf ein gesteigertes Gruppenpotenzial durch hohen individuellen Fertigkeitenzuwachs zurückzuführen sein.

<sup>83</sup> Die statistische explorative Überprüfung der Mediation in Kapitel 6.4.5 deutete darauf hin, dass G-I Transfer als Mediator in Experiment 3 wirksam war. Hierbei wurde als unabhängige Variable die Realbedingung mit den beiden Nicht-Realbedingungen kontrastiert. Dies ist methodisch nicht unproblematisch, da die Leistung in der Nominalbedingung sich im Einzelvergleich nicht von der Leistung in der Realbedingung unterschied. Statistisch sauber könnte die Mediation in Experiment 3 nur unter Ausschluss der Nominalbedingung nachgewiesen werden.

Ein zweites methodisches Problem für Experiment 3 entstand daraus, dass der Mediator G-I Transfer erst nach der Erhebung der abhängigen Variablen Gruppenleistung gemessen werden konnte und das Differenzmaß für G-I Transfer zwischen den individuellen Steuerungsphasen 1 und 3 auch I-I Transfer nach der Gruppeninteraktion beinhaltete, dessen Ausmaß möglicherweise zwischen den Bedingungen variierte und nicht eindeutig bestimmbar war.

Das in den ersten beiden Experimenten verfügbare Differenzmaß der individuellen Steuerungsfertigkeit zwischen den Phasen 1 und 2 beinhaltete durch die alternierenden Gruppenurteile in Phase 2 resultierende begünstigende Systemausgangszustände im Vergleich zur Individualbedingung und war somit gänzlich ungeeignet um G-I Transfer als Mediator zu untersuchen.

Nominalgruppen mit gemitteltem Systemverlauf hinausgehen. Auf Ebene der Einzelstudien war der Leistungsvorteil der real interagierenden Gruppen gegenüber dem statistischen Aggregationsvorteil nur deskriptiv nachweisbar und manifestierte sich bei der Optimierungsaufgabe in Experiment 1 und 3 systembedingt erst in der zweiten Hälfte des Systemverlaufes. Ob eine Realgruppenleistung den Nominalgruppenleistungen überlegen ist, sollte nicht nur unreflektiert durch eine Betrachtung der statistischen Signifikanz bzw. der  $p$ -Werte erfolgen, sondern immer im Hinblick auf die Restriktionen des Systems und der Steuerungsbedingungen erfolgen. Wenn Steuerungserfolge sich systembedingt erst nach einiger Zeit manifestieren sind auch resultierende Effektgrößen davon abhängig, wie viele Steuerungstrials den Probanden zur Verfügung gestellt werden. Das Leistungsniveau der Realgruppen bewegte sich in allen drei Experimenten zwischen resultierender Leistung bei optimaler Steuerung unter der jeweiligen optimal rule und bei unbeeinflusster Entwicklung der Systemzustände unter der no control rule. Deskriptiv lagen die Realgruppen im Vergleich zu den anderen Bedingungen näher am Leistungsniveau der optimalen Steuerung. In dem komplexeren Wirtschaftsszenario des Experimentes 2 entsprach die durchschnittliche individuelle Steuerungsleistung nur dem Systemverlauf unter der no control rule. Daraus lässt sich schließen, dass der Einsatz von Gruppen insbesondere bei höherer Systemkomplexität gegenüber dem Einsatz individuell steuernder Personen bevorzugt werden sollte. Der Leistungsvorteil von Gruppen muss allerdings mit einer signifikant längeren Steuerungsdauer und in der Praxis einhergehenden höheren Personalkosten bezahlt werden.

Die Systemsteuerungsleistung der interagierenden Gruppen ließ jedoch noch viel Verbesserungsspielraum. Die theoretisch resultierende Steuerungsleistung bei Umsetzung des pro Trial akkuratesten Urteils bzw. die konsistente Auswahl der Individualurteile des besten Gruppenmitglieds während der Steuerungsphase 2 hätten die Gruppenleistung signifikant steigern können. Die Gruppen erzielten jedoch keine Koordinationsgewinne. Die Gruppenleistung entsprach im gering komplexen Kühlsystemszenario in Experiment 1 lediglich der mittleren individuellen Leistung respektive dem Gruppenpotenzial. Im ungleich komplexeren Wirtschaftsszenario in Experiment 2 traten sogar Koordinationsverluste auf und die Gruppenleistung fiel geringer aus als die potenzielle Gruppenleistung der Mitglieder.

Als mögliche Ursache für die ausbleibenden Koordinationsgewinne wurde eine Gewichtung nach Extraversion, wie sie bei statischen Urteilsaufgaben gefunden wurde (vgl. Bonner, Sillito, & Baumann, 2007), ausgeschlossen. Auch die bewusste Anwendung einer Mittelungsstrategie traf gemäß der retrospektiv geäußerten Gewichtungsententionen der

Gruppenmitglieder nicht zu<sup>84</sup>. Die Gruppenmitglieder beabsichtigten das beste (Experiment 1) bzw. das beste und mittlere Gruppenmitglied (Experiment 2) stärker zu gewichten. Sie erkannten auch überzufällig häufig das gemäß der Individualleistung in Phase 1 beste Gruppenmitglied. Allerdings änderte sich während der Gruppensteuerungsphase in mehr als 50% der Fälle die Kompetenzrangreihe innerhalb der Gruppen, sodass die subjektive Kompetenzgewichtung in Bezug auf die Gewichtung nach tatsächlicher Kompetenz in einer Mittelungsstrategie resultierte. Diese Befunde bei den untersuchten dynamischen Systemszenarien legen die Schlussfolgerung nahe, dass Leistungsfeedback gemäß der vor der Gruppeninteraktion vorliegenden Mitgliederkompetenz keine geeignete Intervention darstellt um die Gruppenleistung zu steigern.<sup>85</sup>

## **7.2 Leistungs- und wissensbezogener Group-to-Individual Transfer bei der Steuerung dynamischer Systeme**

Die Ergebnisse der Experimente 1 und 3 mit dem Kühlsystemszenario zeigten, dass die Qualität des individuellen Systemwissens (gemessen über die Anzahl richtiger, geäußerter Systemannahmen) nach der Gruppeninteraktion signifikant höher ausfiel als nach individueller Systemsteuerung und deskriptiv höher als nach der Steuerung in Nominalgruppen mit gemitteltem Systemverlauf. Der Gruppenkontext bietet somit im Rahmen eines gering komplexen dynamischen Systems Möglichkeiten, falsche Systemannahmen zu korrigieren und neue richtige Systemannahmen zu erwerben, die über die Selbsterfahrung bei individueller Übung oder die Kenntnisnahme von Gruppenmittelwerten als Bezugsrahmen hinausgehen. Es fanden sich Hinweise darauf, dass der wissensbezogene G-I Transfer als Mediator für die Gruppenleistung wirksam sein könnte. Im vergleichsweise komplexen Wirtschaftsszenario des Experimentes 2 konnte hingegen kein signifikant größerer Wissenszuwachs der Realgruppenmitglieder nach der Gruppeninteraktion festgestellt werden. Dieses Ergebnis steht in Einklang mit vorherigen Befunden. Beispielsweise hat Broadbent (1977) festgestellt, dass Probanden, die individuell die Steuerung eines dynamischen Verkehrssystems übten, anschließend nicht

---

<sup>84</sup> Trotz der subjektiv geäußerten differenzierten Gewichtungsententionen nach Mitgliederexpertise ist es nicht auszuschließen, dass die Probanden während der Steuerung eine Mittelungsstrategie anwandten und erst retrospektiv im Fragebogen eine Gewichtung nach Mitgliederexpertise vorgaben, um zu ihren vorherigen Angaben der wahrgenommen Kompetenzunterschiede zwischen den Gruppenmitgliedern konsistent zu antworten.

<sup>85</sup> Leistungsfeedback könnte dennoch einen leistungsförderlichen Effekt haben, falls es allgemeines Reflektieren über die Leistung der Gruppenmitglieder anregt. Dies sollte in zukünftiger Forschung empirisch überprüft werden.

besser explizite Wissensfragen zu den Systemzusammenhängen beantworten konnten. Die Steuerungserfahrung steigerte jedoch das implizite Wissen, das Broadbent über die Anzahl sinnvoller Systementscheidungen erfasste. So ist es auch für das Szenario im zweiten Experiment denkbar, dass die Gruppeninteraktion sich (über die reine Übung in der Individualbedingung hinaus) positiv auf die implizite Steuerungsfertigkeit der Gruppenmitglieder auswirkte. In welchem Ausmaß in Experiment 2 G-I Transfer vorlag, konnte nicht eindeutig bestimmt werden, da die individuellen Steuerungsurteile alternierend zu den Gruppenurteilen erhoben wurden und so die gruppenurteilsbasierten günstigen Systemzustände in die Individualleistung einfließen.

In Experiment 3 konnte durch ein erweitertes Design mit einer individuellen Steuerungsphase nach der Gruppenphase als ein zentraler Befund dieser Arbeit nachgewiesen werden, dass stabile individuelle Fertigungsgewinne durch die gemeinsame Steuerung in Realgruppen auftraten. Die Gruppeninteraktion ging mit individueller Fertigungsverbesserung einher, die signifikant größer als die Übungseffekte von individuell steuernden Probanden oder Nominalgruppenmitgliedern, die gemittelte Referenzwertinformationen erhielten, ausfiel. Die Gruppeninteraktion wirkt sich also nicht nur positiv auf die erzielte Steuerungsleistung aus, sondern hat (zumindest für einen Teil der Gruppenmitglieder) auch langfristige positive Auswirkungen. Der G-I Transfer wurde differenziert auf Ebene des einzelnen Gruppenmitglieds betrachtet. Das anfangs schlechteste und mittlere Gruppenmitglied näherten sich dem Leistungsniveau des vor der Steuerung besten Gruppenmitglieds an. Letzteres stagnierte auf dem Ausgangsniveau oder verschlechterte sich deskriptiv leicht während bzw. nach der Gruppeninteraktion. In weiterführenden Studien könnte untersucht werden, ob Leistungsfeedback bzgl. der „guten“ individuellen Ausgangsleistung für das beste Gruppenmitglied sich förderlich auf dessen G-I Transfer auswirken würde, da es sich weniger stark durch gegensätzliche Steuerungsstrategien (insbesondere zu Beginn der Gruppenphase) der schlechteren Gruppenmitglieder beeinflussen lassen sollte. Gleichzeitig sollte aber auch betont werden, dass die Kompetenzrangreihe sich während der Gruppenphase verändern kann. Das schlechteste Gruppenmitglied profitierte mit großem G-I Transfer von der Gruppeninteraktion. Dieses Ergebnis ist in Einklang mit dem Ergebnis von Schultze et al. (2012) bei statischen Schätzaufgaben und widerspricht dem Ergebnis von Bonner und Baumann (2008), die festgestellt hatte, dass sich die Leistung des schlechtesten Mitglieds noch weiter verschlechtern kann, wenn dieses mit zu vielen zusätzlichen Informationen konfrontiert wird.

Auffallend war in der Nominalbedingung mit gemitteltem Systemverlauf die signifikante Verschlechterung der Individualleistung des besten Mitglieds in den Experimenten 2 und 3. Eine mögliche Erklärung für diese Verschlechterung könnte sein, dass die guten Gruppenmitglieder taktisch denken und versuchen durch extremere Urteile (als sie sie bei individueller Bearbeitung treffen würden) die „schlechteren“ Eingaben der anderen Mitglieder zu kompensieren. Ob dies der Fall ist, könnte in nachfolgenden Untersuchungen durch Selbstauskunft zum taktischen Vorgehen und durch ein weiteres Eingabefeld für die individuellen Urteile der Nominalgruppenmitglieder unter der Annahme, dass die Qualität dieses Urteil für die Bonuszahlung zählt, überprüft werden.

### **7.3 Beschränkungen und Implikationen für die Praxis und weitere Forschung**

Die Ergebnisse dieser Arbeit implizieren, dass der Einsatz von Realgruppen für die Leistungsoptimierung empfehlenswert ist. Zwar zeigen sich in Relation zu Nominalgruppen mit gemitteltem Systemverlauf nur kleine Effekte; diese feinen Unterschiede können aber in der Praxis große Auswirkungen haben. Dem möglichen Gewinn des Unternehmens in Bezug auf Sicherheitsoptimierung oder Kostenersparnis stehen jedoch höhere Personalkosten und ein größerer Zeitaufwand für die Steuerung gegenüber. Es bleibt zu untersuchen, ob Realgruppen bei einer kürzeren zeitlichen Taktung der Steuerungsurteile, die der individuellen Steuerung entspricht, einen gleichbleibenden Leistungsvorteil erzielen. Für Steuerungsaufgaben ohne vorgegebene Taktung und ohne Entscheidungsschnelligkeit als Teilleistungskriterium wird es sich im Praxiskontext wahrscheinlich rentieren, höhere Personalkosten für einige zusätzliche Arbeitsminuten zu investieren, um überproportional größere Systemoptimierungen bzw. Kosteneinsparungen durch Realgruppenurteile zu erzielen. Ob Realgruppen auch Steuerungsvorteile in zeitkritischen Steuerungssituationen mit schneller eigendynamischer Verschlechterung des Systemzustands bei ausbleibenden Eingriffen erzielen können, ist zukünftig zu untersuchen.<sup>86</sup>

Neben dem Gewinn des Steuerungsausgangs für die Organisation, die das dynamische System betreibt, bietet die Steuerung in Gruppen auch positive Effekte für die Mitarbeiter. Die Gruppenmitglieder erzielten individuellen Lernzuwachs in Bezug auf das

---

<sup>86</sup> Ein entsprechendes computersimuliertes Szenario zur kontrollierten laborexperimentellen Untersuchung von interaktionsbedingten Prozessgewinnen unter leistungskritischer Urteilsgeschwindigkeit stellt *Networked Fire Chief* von Omodei und Wearing (1993; 1995b) dar. Hierbei wird eine Feuerlöschsituation simuliert, bei der sich schnell ausbreitende Flächenbrände bekämpft werden müssen.

Systemwissen und die Steuerung, der über Übungseffekte bei Einzelarbeit hinausging. Somit kann die Systemsteuerung das Leistungsmotiv der Mitarbeiter ansprechen (vgl. McClelland, 1964) bzw. das menschliche Grundbedürfnis nach Kompetenz befriedigen (vgl. Deci & Ryan, 1985). Diese Bedürfnisbefriedigung der Mitarbeiter kann mit höherer Arbeitszufriedenheit als bei Einzelsteuerung einhergehen. Eine hohe Arbeitszufriedenheit könnte sich wiederum positiv auf Krankheitstage, Absentismus, das Auftreten innerer Kündigung oder Fluktuation der Belegschaft auswirken. Somit kann auch die Organisation vom höheren Befriedigungspotenzial der Gruppenarbeit profitieren.<sup>87</sup>

Für die Zusammenstellung von besonders leistungsfähigen Gruppen lassen sich aus den Ergebnissen dieser Arbeit keine besonderen Empfehlungen in Bezug auf demographische Eigenschaften der Gruppenmitglieder ableiten. Förderlich für die Gruppenleistung ist eine gute individuelle Vorbildung der Gruppenmitglieder (im Sinne einer hohen Startkompetenz der Gruppe). Wie sich herausgestellt hat, ist Leistungsfeedback in Bezug auf die vorangehende Individualleistung ungeeignet zur Steuerungsoptimierung der Gruppen, da die Kompetenzrangreihe der Gruppenmitglieder häufig nicht zeitstabil ist.

Die in dieser Arbeit verwendeten dynamischen Systeme beinhalteten die Charakteristika nonlinearer Variablenbeziehungen und zeitverzögerter Systemreaktion auf Steuerungseingriffe, die in nahezu allen technischen, biologischen und wirtschaftlichen Systemen in der Praxis vorzufinden sind. Gleichwohl ist die Generalisierbarkeit der Befunde durch die vier Besonderheiten der Gruppengröße, der Verwendung von ad hoc Gruppen, der selbstbestimmten Steuerungsgeschwindigkeit sowie des begrenzten Komplexitätsbereiches der verwendeten Szenarien eingeschränkt. Implikationen dieser Spezifika für die Generalisierbarkeit werden im Folgenden genauer betrachtet.

In dieser Studie wurde ausschließlich die Leistung von Dreiergruppen bei der Steuerung dynamischer Systeme untersucht und mit der Nominalgruppenleistung verglichen. Daher ist offen, inwiefern die Ergebnisse auf andere Gruppengrößen übertragbar sind. Ergebnisse von Studien, die Gruppenleistung in Abhängigkeit von der Gruppengröße bei dynamischer Systemsteuerung (Wolfe & Chacko, 1983) oder statischen Problemlöseaufgaben untersucht haben (Laughlin, Hatch, Silver, & Boh, 2006), kamen konsistent zu dem Ergebnis, dass Dreiergruppen in Relation zu Dyaden bessere Leistungen

---

<sup>87</sup> Zusätzliche, nicht im Hauptteil der Arbeit aufgeführte, explorative Analysen zeigten außerdem, dass das Wohlbefinden während der Gruppenarbeit höher ausfiel als bei Einzelarbeit (vgl. Anhänge F, L und N) und durch die Gruppenarbeit daher auch das Anschlussmotiv (vgl. McClelland, 1964) bzw. das menschliche Grundbedürfnis nach sozialem Bezug (vgl. Deci & Ryan, 1985) angesprochen wird.



erbringen, es aber keinen weiteren signifikanten Leistungszuwachs bei vier oder mehr Gruppenmitgliedern gibt. Entsprechend ist anzunehmen, dass der Leistungsvorteil von Dyaden in Relation zu Dreiergruppen niedriger ausfallen würde und auch das Ausmaß an G-I Transfer aufgrund der kleineren Wissensbasis in der Gruppe geringer ausfällt (was aber noch zu prüfen wäre).

Eine weitere Besonderheit dieser Studie betrifft die Stichprobe, die sich auf Studierende und ad hoc zusammengestellte Gruppen beschränkte. Der Bildungsstand der Stichprobe dürfte dem der meisten Systemoperateure entsprechen, die Abitur und anschließendes Studium oder eine Fachausbildung absolviert haben. Allerdings ist noch zu untersuchen, inwiefern die Ergebnisse auf Fachexperten mit langjähriger Berufserfahrung sowie auf langfristig bestehende Arbeitsteams übertragbar sind. Langfristig bestehende Arbeitsteams, auch als *traditioned work groups* bezeichnet (vgl. Lorge et al., 1956), haben sich über einen Zeitraum gemeinsam entwickelt, zeigen Commitment gegenüber der Gruppe und verfolgen kooperativ organisationale Ziele. Ad hoc Gruppen sind hingegen durch einen Experimentator zusammengestellt und bearbeiten eine spezifische extern festgelegte Aufgabe. Die Gruppenmitglieder haben keine gemeinsame Vorgeschichte und erwarten auch nicht zukünftig zusammenarbeiten, was sich auf den Interaktionsprozess auswirken kann. Denkbar wäre beispielsweise, dass Personen gegenüber anderen Mitgliedern in ad hoc Gruppen zurückhaltender agieren, als es eingespielte Arbeitsgruppen tun und daher eine demokratische Mittelungsstrategie einsetzen, die den gesellschaftlichen Normen entspricht und die Auseinandersetzung mit Kompetenzunterschieden vermeidet. Arbeitsgruppen weisen anders als ad hoc Gruppen darüber hinaus häufig Hierarchieunterschiede auf und beinhalten eine Führungskraft. Diese Führungskraft lenkt und dominiert in der Regel den Informationsaustausch und die Entscheidungsfindung in der Gruppe, was sich in Abhängigkeit der konkreten Interaktionsgestaltung leistungsförderlich oder hinderlich auf den Gruppenprozess auswirken kann (vgl. Cruz, Henningsen, & Smith, 1999, für eine ausführliche Diskussion). Außerdem ist es wahrscheinlich, dass sich in langfristig bestehenden Arbeitsgruppen ein *transaktives Wissenssystem* (Wegner, 1986) herausbildet, das Metawissen darüber beinhaltet, wer in der Gruppe was weiß und Experte ist. Zhang, Hempel, Han und Tjosvold (2007) konnten in Arbeitsteams organisationsübergreifend zeigen, dass ein ausgeprägteres transaktives Wissenssystem mit Steigerungen der Arbeitsleistung einhergeht. Transaktive Wissenssysteme sind dann besonders hilfreich, wenn Gruppen mit teilbaren Aufgaben konfrontiert sind, deren Teilaufgaben von den Gruppenmitgliedern mit dem jeweiligen Spezialwissen bearbeitet werden. Aber auch für unteilbare Systemsteuerungsaufgaben, bei denen alle

Gruppenmitglieder gemeinsam Steuerungsurteile treffen, ergeben sich in der Praxis Vorteile durch ein transaktives Wissenssystem. So könnten sich Gruppen bei besonderen Situationen wie Störfällen in Kernkraftwerken stärker an den Urteilen der Gruppenmitglieder orientieren, die in der Vergangenheit Erfahrungen mit Zwischenfällen gemacht haben. Auch sollten eingespielte Arbeitsteams mit hoher Wahrscheinlichkeit den Steuerungsexperten mit der besten Individualleistung identifizieren können, was den ad hoc Gruppen in dieser Studie nicht überzufällig gut gelang. In weiteren Studien, die ad hoc Gruppen und eingespielte, natürliche Arbeitsgruppen vergleichen, sollten mögliche Unterschiede im Interaktionsprozess der Systemsteuerung systematisch anhand von Videoanalysen der Gruppeninteraktion untersucht werden.

Ein dritter besonderer Aspekt, der für alle Experimente dieser Studie gilt, ist die selbst gewählte zeitliche Taktung der Steuerungseingriffe und die daran gekoppelte Systemreaktion. Diese Art der Systemsteuerung betrifft in der Realität einige Steuerungskontexte wie die Produktionssteuerung. Andere Steuerungskontexte erfordern hingegen Steuerungsurteile in Echtzeit, deren Taktung durch situative Erfordernisse vorgegeben wird. So müssen beispielsweise Fluglotsen bei der Luftraumkontrolle die Fluggeschwindigkeiten vieler Maschinen bei der Vergabe von Landeerlaubnissen berücksichtigen (vgl. Gonzalez, Vanyukov, & Martin, 2005). Bei ungeplanten Zwischenfällen, wie einer erforderlichen Notlandung eines Flugzeuges oder Hindernissen im Luftraum, haben sie nur begrenzt Zeit Flugzeuge zielgerichtet umzuleiten, um Kollisionen und Unfälle zu vermeiden. Die Realgruppen in dieser Studie haben ihren Leistungsvorteil gegenüber der individuellen Steuerung bei signifikant längerer Steuerungsdauer erzielt. In zukünftigen Untersuchungen sollte getestet werden, ob die Überlegenheit von Realgruppen auch bei Systemsteuerung in Echtzeit mit kürzeren Urteilsintervallen erhalten bleibt. Wie bereits angemerkt, bietet sich für die laborexperimentelle Untersuchung von interaktionsbedingten Prozessgewinnen unter leistungskritischer Urteilsgeschwindigkeit beispielsweise eine Feuerwehreinsatzsimulation zur Bekämpfung von Großflächenbränden an (vgl. Omodei & Wearing, 1993).

Die vierte Einschränkung der Ergebnisse dieser Arbeit im Hinblick auf die Generalisierbarkeit stellt der Komplexitätsbereich der verwendeten Steuerungsszenarien dar. Obgleich die Komplexität variiert wurde und im Wirtschaftsszenario des zweiten Experimentes mit einer größeren Anzahl an Variablen(-beziehungen) höher ausfiel als im Kühlsystemszenario des Atomkraftwerkes, ist unklar, inwiefern die Ergebnisse auf hoch komplexe Szenarien wie die Steuerung der Stadt Lohhausen mit 2000 Systemvariablen (Dörner et al., 1983) übertragbar sind. Die zukünftige Überprüfung der Replizierbarkeit der

Befunde zu Prozessgewinnen und G-I Transfer mit dem grundlegenden experimentellen Design dieser Arbeit wäre für diese extrem komplexen Steuerungskontexte wünschenswert.

In dieser Arbeit wurde festgestellt, dass die Interaktion in Realgruppen zu stabilem G-I Transfer führt, der über den Lerntransfer in der Nominalbedingung und Individualbedingung hinausgeht. Für die weitere Forschung ist es eine interessante Frage, welche quantitativen und qualitativen Interaktionsprozesse für das Auftreten von G-I Transfer notwendig bzw. besonders förderlich sind. Hierzu wäre eine detaillierte Videokodierung der Inhaltsaspekte des Interaktionsprozesses erforderlich. „Der Vorgang der Kodierung besteht zunächst darin den Inhaltsstrom sozialer Interaktion in einzelne sinntragende Einheiten (Interakte) zu zerlegen“ (Boos, Scharpf, & Fisch, 1991, S. 116 f.). Anschließend sollten die Interakte nach Inhaltskategorien sortiert und ausgewertet werden. Als übergeordnete Kategorien bieten sich nach Boos et al. (1991) *Problemstrukturierung*, *Lösungssuche* und *Entscheidungsfindung* an. Die Problemstrukturierung würde in Bezug auf das dynamische System Äußerungen zu Systemeigenschaften wie zeitverzögerter Reaktion und Variablenzusammenhängen beinhalten. Hierbei ist entscheidend, inwiefern individuelle Sichtweisen zu einer komplexen Gesamtrepräsentation verknüpft werden und inwiefern eingebrachte Sichtweisen von den anderen Gruppenmitgliedern als zutreffend erkannt werden. In die Kategorie Lösungssuche fallen Vorschläge zu zielführenden Steuerungsstrategien. Die Kategorie Entscheidungsfindung umfasst das Abwägen von Argumenten, die für oder gegen eine bestimmte Strategie sprechen und die Einigung auf Gruppenurteile zur Steuerung. Hierbei könnte die bisher nur indirekt aus den Daten (bzw. post hoc und pauschal für die Gesamtsteuerung) erfragte Gewichtung der Einzelurteile für jeden einzelnen Trial kodiert werden. Anschließend könnte der prozentuale Anteil der Trials bestimmt werden, in dem bestimmte Urteilsschemata wie Mittelwertbildung oder Übernahme des besten Individualurteils zur Anwendung kamen. Darüber hinaus wäre es interessant zu untersuchen, ob bestimmte Gesprächstaktiken zur Einflussnahme auf die Gruppenurteile eingesetzt werden (vgl. Littlepage, Schmidt, Whisler, & Frost, 1995), und ggf. inwiefern die eingesetzten Argumentationstaktiken sich auf die Kompetenzwahrnehmung und die Beitragsgewichtung für die Gruppenurteile auswirken.

Die Auswertung zeigte, dass es in dieser Studie keinen Zusammenhang von Extraversion und Expertisezuschreibung sowie Gewichtung gab. Dabei wurde implizit auf Basis vorangehender Studien angenommen, dass Personen mit höherer Extraversionsausprägung einen höheren Redeanteil hätten und aufgrund ihrer sozialen Fertigkeiten als kompetenter wahrgenommen würden. Anhand der Interaktionsanalyse

könnte empirisch überprüft werden, ob der Redeanteil der Gruppenmitglieder bei der Systemsteuerung tatsächlich mit Extraversion korreliert oder aber Zusammenhänge mit Studienfach, Argumentationstaktiken oder anderen Faktoren eine Rolle spielen.

Auch für die Interaktionsprozessanalyse bietet es sich an, besonders erfolgreiche und erfolglose Gruppen zu kontrastieren. So könnte analysiert werden, ob ein bestimmter Phasenverlauf für die Spitzenleistungsgruppen charakteristisch ist und dort z.B. anfangs ein erschöpfender Wissensaustausch stattfindet, bevor mit der Steuerung begonnen wird. Aus dem Vorgehen der Spitzenleistungsgruppen könnten dann generelle Empfehlungen abgeleitet und deren Wirksamkeit in weiteren Studien getestet werden.

Die Ergebnisse von Schultze et al. (2012) deuten darauf hin, dass schon eine einmalige Gruppeninteraktion bei Urteilsaufgaben G-I Transfer hervorrufen kann. Meyer und Scholl (2009) konnten für die dynamische Systemsteuerung zeigen, dass eine unstrukturierte Diskussion über das System (ohne gemeinsame Steuerung in der Gruppe) die individuelle Steuerungsleistung steigern kann. In zukünftigen Studien wäre es wichtig zu untersuchen, welche Interaktionsdauer und Art der Interaktion notwendig und ausreichend sind, um G-I Transfer bei den Gruppenmitgliedern zu erzielen.

Ein wichtiges Anliegen zukünftiger Forschung sollte die Untersuchung der Frage sein, wie man Gruppenleistung und interaktionsbedingten Lerntransfer fördern kann, um noch deutlichere Prozessgewinne und größere individuelle Fertigungsgewinne zu erzielen und dadurch die Steuerung dynamischer Systeme sicherer und effizienter zu machen. Aus der Sichtung der bisherigen Forschung zur Gruppenleistung bei anderen Aufgabentypen wie quantitativen Schätzaufgaben (z.B. Bonner & Baumann, 2012), Auswahlentscheidungen (z.B. Schulz-Hardt, Brodbeck, Mojzisch, Kerschreiter, & Frey, 2006), strategischen Entscheidungsaufgaben (z.B. Schweiger, Sandberg, & Rechner, 1989) und statischen komplexen Problemlöseaufgaben (z.B. Bales & Strodtbeck, 1951), aus der Forschung zu dynamischen Systemsteuerungsaufgaben (z.B. Reither, 1979) und aus den Ergebnissen dieser Arbeit lassen sich meines Erachtens fünf Ansätze<sup>88</sup> zur Förderung der Gruppenleistung ableiten, die auch miteinander kombiniert werden können:

---

<sup>88</sup> Ein weiterer, theoretisch geeigneter Ansatz ist das Bereitstellen von individuellem Leistungsfeedback vor der Gruppeninteraktion, um der Gruppe die relative Steuerungskompetenz der Gruppenmitglieder transparent zu machen und eine leistungsabhängige Gewichtung der Individualurteile für das Gruppenurteil sowie G-I Transfer zu erleichtern. In dieser Arbeit erwies sich Leistungsfeedback jedoch überwiegend als ungeeignet, da die Leistungsrangreihe innerhalb der Gruppen zwischen der vorangehenden individuellen Arbeitsphase und der Gruppenphase nicht

- *Nutzung vorhandenen Wissens fördern*
- Instruktionen zur *akkuratheitsbasierten Gewichtung* geben<sup>89</sup>
- kontinuierliche *Selbstreflexion* des Wissensstandes und der gewählten Steuerungsstrategien anregen
- allgemeine Instruktionen zur *Strukturierung des Interaktionsprozesses erteilen*
- *Dissens* durch Gruppenzusammensetzung und/oder spezielle Gesprächstechniken fördern.

Dass Interventionen zur besseren Nutzung schon vorhandenen Wissens den Wissenstransfer, die Demonstrierbarkeit richtiger Beiträge, das Erkennen von Expertise, die stärkere Gewichtung guter Beiträge und somit die Gruppenleistung steigern können, haben Bonner und Baumann (2012) gezeigt. In ihrer Untersuchung lösten Gruppen quantitative Schätzaufgaben. Die Gruppenmitglieder erhielten die Instruktion, vor der Abgabe von Individualurteilen ihr individuelles Wissen zu aktivieren und zwei Wissensinhalte zu notieren, die mit der gestellten Aufgabe in Zusammenhang stehen. Anschließend bearbeiteten die Probanden dieselben Aufgaben in der Gruppe. Die Leistung der Gruppen mit der Wissenstransfer-Instruktion war sowohl der Nominalgruppenleistung als auch der Leistung von Gruppen ohne Transfer-Instruktion überlegen. Gleichzeitig war die Gruppeninteraktion unter der Wissenstransfer-Instruktion zielführender: Bei gleicher Diskussionsquantität wurden mehr Wissensinhalte und mehr Informationen über Mitgliederexpertise ausgetauscht als in den Vergleichsbedingungen. Diese erste Interventionsinstruktion zum Heranziehen vorhandenen Wissens aus analogen Problemsituationen zur Leistungssteigerung bei der dynamischen Systemsteuerung ist mit wenig Zeitaufwand und geringen Kosten verbunden.

Eine zweite „einfache“ Intervention, deren Effektivität zu testen wäre, ist die Anweisung, bei jedem Steuerungstrial das akkurateste Einzelurteil herauszufinden und als Gruppenurteil auszuwählen. Diese Intervention leitet sich aus dieser Studie ab, die gezeigt hat, dass das Accuracy Model eine signifikante Leistungssteigerung gegenüber der Realgruppenleistung ermöglicht hätte. Hierbei sollte der leistungsförderliche Effekt nicht darauf basieren, dass den Gruppenmitgliedern genau mitgeteilt wird, wie sie in einem konkreten Trial akkuratere Urteile von weniger akkuraten unterscheiden. Solche allgemeingültigen Faustregeln sind für die dynamische Systemsteuerung nicht formulierbar. Die Gruppenmitglieder sollten aber durch die Aufforderung, die individuellen Beiträge zu

---

ausreichend stabil war, um veridikales Feedback in Bezug auf das beste Gruppenmitglied (in beiden Szenarien) und in Bezug auf das schlechteste Gruppenmitglied (im Wirtschaftsszenario) zu geben.

jedem Steuerungszeitpunkt erneut bezüglich der Akkuratheit zu bewerten und für die Gruppenurteilsfindung zu gewichten, darin bestärkt werden kontinuierlich für jeden Trial im Steuerungsverlauf zu hinterfragen, welches das beste Individualurteil sein könnte. Somit soll durch diese Intervention die Reflexion der gewählten Steuerungsstrategie auf Gruppenebene gefördert werden und der pauschale Einsatz von zwei möglicherweise suboptimalen Strategien vermieden werden. Zum einen könnte diese Instruktion der implizit vorhandenen, an demokratische Grundprinzipien angelehnten Mittelung der Individualurteile und Vermeidung der Diskontierung schlechter Einzelbeiträge entgegenwirken und die Probanden ermutigen, explizit kompetente Einzelbeiträge zu identifizieren und umzusetzen. Zum anderen würde durch die Trial-weise Gewichtung nach wahrgenommener Qualität der Individualurteile verhindert, dass die Gruppen über den gesamten Steuerungsverlauf die Vorschläge des (anfangs) subjektiv besten Mitglieds in der Gruppe übernehmen. Wie in dieser Arbeit gezeigt wurde, verändert sich während der Gruppeninteraktion in vielen Fällen die Leistungsreihenfolge innerhalb der Gruppen, sodass eine im Steuerungsverlauf angepasste Gewichtung der Individualurteile (so lange sie überwiegend korrekt erfolgt) einen Leistungsvorteil gegenüber einer unreflektierten, an subjektiver Ausgangsexpertise orientierten Gewichtung bieten sollte. Um den Gruppenmitgliedern das Hinterfragen der Urteilsqualität zu erleichtern, sind darüber hinaus weitere Interventionstechniken denkbar.

Die dritte empfehlenswerte Intervention kann aus Studien zur individuellen Systemsteuerungsleistung abgeleitet werden. Beispielsweise konnte Reither (1979) zeigen, dass kritische Selbstreflexion des eigenen Vorgehens sich leistungssteigernd auswirken kann. Auch Gruppenmitglieder könnten instruiert werden nach Abgabe jedes Individualurteiles im al-G Design kritisch zu hinterfragen, ob sie ihr gesamtes Wissen für die Urteilsabgabe einbezogen haben und ob es nicht andere bzw. bessere Lösungen gäbe. Dieses gezielte Nachdenken könnte zu einer Revision der Individualurteile führen und/oder sich förderlich auf die Diskussion der Individualurteile in der Gruppe und die Korrektheit der Gruppenurteile auswirken.

Als vierte leistungsförderliche Intervention bietet sich die Strukturierung des Interaktionsprozesses in der Gruppe an. Schaub und Strohschneider (1992) stellten fest, dass erfolgreiche individuelle Systemsteuerer in einem Entwicklungshilfeszenario sich dadurch auszeichneten, dass eine Phase der Informationssammlung über das System dem Fällen von Steuerungsurteilen vorausging. Daher scheint auch für Gruppen die explizite Instruktion empfehlenswert, den Interaktionsprozess in eine Orientierungsphase mit systematischer Informationssammlung, eine Evaluationsphase mit der systematischen

Bewertung von Alternativen und eine Kontrollphase mit der Steuerungsausführung zu gliedern (vgl. zur Terminologie Bales & Strodtbeck, 1951).<sup>90</sup> In der Bewertungsphase ist eine kritische Betrachtung des geäußerten individuellen Systemwissens und der vorgeschlagenen Steuerungsurteile wichtig, um dem Treffen inadäquater Entscheidungen durch das von Janis (1982) beschriebene *Groupthink* Phänomen vorzubeugen.<sup>91</sup>

Die fünfte und die sechste mögliche Intervention zur Förderung der kritischen Auseinandersetzung mit Systemverständnis und Steuerungsstrategien beziehen sich auf das gezielte Herbeiführen von Dissens durch die Gruppenzusammenstellung bzw. durch den Einsatz Dissens förderlicher Gesprächstechniken. Eine Studie von Schulz-Hardt, Brodbeck, Mojzisch, Kerschreiter und Frey (2006) zeigt, dass anfänglicher Dissens der Gruppenmitglieder bei Entscheidungsaufgaben die Diskussionsintensität erhöht und die Entscheidungsqualität steigert. Dissens wurde dabei über unterschiedliche Entscheidungspräferenzen innerhalb der Gruppe operationalisiert. Daraus lässt sich für Steuerungsgruppen eine möglicherweise leistungsförderliche Gruppenzusammenstellung ableiten, bei der das individuelle (zuvor erfragte) Systemverständnis der Gruppenmitglieder unterschiedliche und konträre Inhalte aufweist. Dies sollte ebenfalls die Diskussion anregen, die Identifikation richtiger Wissensinhalte und besserer Strategien fördern und dadurch zu Leistungssteigerungen beitragen. Kurzfristig wäre eine Verschlechterung der Gruppenleistung denkbar, falls die Richtigkeit unterschiedlicher Sichtweisen sukzessiv durch Anwendung über Steuerungsabschnitte überprüft wird. Unter welchen Umständen Realgruppen Dissens bezüglich dynamischer Systeme ausdiskutieren oder aber Steuerungsstrategien durch Ausprobieren testen, muss in weiteren Studien untersucht werden. Bei geringer verbaler Demonstrierbarkeit der Qualität von Steuerungsvorschlägen bzw. Strategien bezüglich der Urteileingabe über längere Sequenzen von Trials wäre das praktische Testen besonders wahrscheinlich.

Um negative Auswirkungen des vorschnellen Strategietestens auf die Leistung zu reduzieren, können Interventionen eingesetzt werden, die Dissens bereits im

---

<sup>90</sup> Darüber hinaus wird es in den meisten Fällen erforderlich sein, während der Steuerungsphase kontinuierlich das Systemfeedback im Hinblick auf den Erfolg der gewählten Eingriffsstrategie zu evaluieren und gegebenenfalls Anpassungen vorzunehmen.

<sup>91</sup> Gruppendenken ist nach Janis (1982) ein Denkmodus in nach außen isolierten, kohäsiven Gruppen, der zu starkem Konformitätsdruck führt und unter anderem individuelle Kritik der Gruppenmitglieder an irrationalen, suboptimalen Entscheidungen unterbindet. Dadurch können negative Nebenwirkungen übersehen oder ausgeblendet werden und die Gruppenleistung fällt schlechter aus als unter systematischer Abwägung aller bekannten Handlungsalternativen möglich gewesen wäre.

vorangehenden Diskussionprozess fokussieren. Hierunter fallen *Devil's Advocacy* (DA) und *Intellectual Inquiry* (DI; vgl. Cosier, 1978; Schweiger, Sandberg, & Rechner, 1989). Beide Techniken können auch als spezielle (da auf das Hervorrufen von Dissens abzielende) Strukturierungen des Interaktionsprozesses interpretiert werden. Sie haben sich als effektiv erwiesen, um die Qualität strategischer Entscheidungen bei der Bearbeitung von nicht dynamischen strategischen Case Studies in Gruppen zu steigern und es liegen Hinweise auf eine Steigerung der individuellen Steuerungsleistung bei einem komplexen dynamischen Problem unter Anwendung der Devil's Advocacy Technik (DA) vor (Cosier & Rechner, 1985).

Bei der Devil's Advocacy Technik wird die Gesamtgruppe in zwei Gruppen geteilt, wobei eine Subgruppe Vorschläge macht und durch Argumente stützt und die andere Subgruppe (oder ein Gruppenmitglied) als Teufelsanwalt Kritik übt, woraufhin die erste Subgruppe ihre Vorschläge revidiert bis eine annehmbare Lösung erreicht ist. Bei der Dialectical Inquiry Technik stellt eine Subgruppe ihre Lösung vor, die zweite Subgruppe überlegt sich einen gegensätzlichen Ansatz und anschließend erarbeitet die Gesamtgruppe eine Konsenslösung. Diese Techniken lassen sich prinzipiell auf dynamische Systemsteuerungsgruppen anwenden. Dabei bietet es sich an, eine der Techniken vor Beginn der Gruppensteuerung in einer Austausch- bzw. Gruppendiskussionsphase einzusetzen und auch während des Steuerungsprozesses die Rolle des Teufelsanwalts zu vergeben, sodass die gewählte Strategie kontinuierlich hinterfragt wird.

Cosier und Rechner (1985) verglichen die individuelle Steuerungsleistung von Studierenden in einer dynamischen Wirtschaftssimulation mit 8 Inputvariablen unter Anwendung von DA, DI, einem „Expertenadvice“, der nur eine Position der DI beinhaltete, und in einer Kontrollbedingung. In der DI und DA Bedingung waren die Subgruppenstellungen bzw. die Teufelsanwaltskritik anhand schriftlich standardisierten Materials vorgegeben. Individuen in der DA Bedingung erzielten dabei signifikant bessere Steuerungsleistungen, wenn die vom Teufelsanwalt kritisierten Steuerungsstrategien inhaltlich riskant waren. Unter der DI Bedingung zeigten sich hingegen keine Effekte der parallel dargestellten risikoreicheren und konservativeren Steuerungsstrategien. Die Effektivität der Dissens-stiftenden Interventionen bei der Steuerung in Realgruppen, wo Strategien und Kritik nicht vorgegeben, sondern von den Gruppenmitgliedern selbst generiert werden, sollte Gegenstand zukünftiger Forschung sein. Am Computer simulierte dynamische Systeme bieten dabei einen sicheren Rahmen um positive Effekte und eventuell auftretende negative Nebeneffekte wie die Beeinträchtigung der Stimmung und Motivation durch die konflikthafte Arbeitsweise zu



untersuchen. In einem zweiten Schritt sollten dann die im Labor getesteten Interventionen, die sich positiv auf den G-I Transfer und die Gruppenleistung auswirken, in der Praxis erprobt werden.

#### **7.4 Abschließende Bemerkung**

Diese Arbeit war durch zwei Hauptanliegen motiviert. Zum einen sollte erstmals mittels der Zusammenführung aktueller sozialpsychologischer Forschungsmethoden und der dynamischen Systemsteuerungsforschung methodisch sauber untersucht werden, ob sich Gruppenarbeit bei diesem Aufgabentyp lohnt, welche Prozesse der Gruppenleistung zu Grunde liegen und welche Konsequenzen sich für den Wissenserwerb und die resultierende Leistung ergeben. Zum anderen sollte diese Arbeit leistungsförderliche und hinderliche Faktoren aufdecken und eine Anregung für weitere Forschung zu leistungssteigernden Interventionen geben, um die Steuerung von technischen, biologisch-ökologischen und wirtschaftlichen Systemen in der Realität sicherer und gewinnbringender zu gestalten. Die Ergebnisse der drei durchgeführten Experimente liefern zwei zentrale Erkenntnisgewinne: Erstens ist der Einsatz von Realgruppen empfehlenswert, da die gruppenbasierte Steuerung die durchschnittliche Leistung von Individuen und von nicht-interagierenden Nominalgruppen übertrifft. Zweitens spricht die starke durchschnittliche Verbesserung individueller Steuerungsfertigkeiten und der Wissensqualität der Realgruppenmitglieder für den Einsatz von Gruppen bzw. Steuerungsteams in der Praxis.

Während der Anfertigung dieser Arbeit ereignete sich die Nuklearkatastrophe in Fukushima, Japan. Dieser Zwischenfall zeigt, dass es trotz vielfältiger sicherheitstechnischer Verbesserungen und Automatisierungen seit der Tschernobylkatastrophe in den 1980ern immer noch zu Situationen kommen kann, in denen die menschliche Steuerungsleistung in technischen Systemen entscheidend für Gesundheit und Leben ist. Auch „Steuerungseingriffe“ von Politikern und Managern in dynamische Finanz-, Wirtschafts- und ökologische Systeme werden zukünftig weit reichende Konsequenzen haben. Entsprechend ist es wünschenswert, dass die Verantwortung vermehrt auf Gruppen von Entscheidungsträgern übertragen wird und die Wirksamkeit der in diesem Teilkapitel angeregten Interventionen zur Leistungssteigerung gruppenbasierter Steuerung in den Fokus zukünftiger Forschung rücken.



## Literatur

- Atwood, M. E. & Polson, P.G. (1976). A process model for water jug problems. *Cognitive Psychology*, 8(2), 191-216.
- Badke-Schaub, P. (1993). *Gruppen und komplexe Probleme: Strategien von Kleingruppen bei der Bearbeitung einer simulierten AIDS-Ausbreitung*. Frankfurt a.M.: Peter Lang.
- Bakken, B. E. (1993). *Learning and transfer of understanding in dynamic decision environments*. Massachusetts, MA: Unveröffentlichte Dissertation.
- Bales, R. F. & Strodtbeck, F. L. (1951). Phases in group problem-solving. *The Journal of Abnormal and Social Psychology*, 46(4), 485-495.
- Barlas Y. & Dalkiran, E. (2008). Testing dynamic decision making under real-time pressure: A scuba diving simulator. *Proceedings of 26th International system dynamics conference*. (CD). Albany, NY, USA, July 2008. Retrieved online from: <http://www.systemdynamics.org/conferences/2008/proceed/papers/>
- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology*, 51(6), 1173-1182.
- Berkowitz, M. S. & Szabo, M. (1978). Individual and dyadic problem solving as a function of mental ability through computer simulation. *Journal of Educational Research*, 71, 328-333.
- Berry, D. C. & Broadbent, D. E. (1984). On the relationship between task performance and associated verbalizable knowledge. *Quarterly Journal of Experimental Psychology*, 36A, 209-231.
- Bonner, B. L. & Baumann, M. R. (2008). Informational intra-group influence: The effects of time pressure and group size. *European Journal of Social Psychology*, 38(1), 46-66.
- Bonner, B. L. & Baumann, M. R. (2012). Leveraging member expertise to improve knowledge transfer and demonstrability in groups. *Journal of Personality and Social Psychology*, 102(2), 337-350.
- Bonner, B. L. (2000). The effects of extroversion on influence in ambiguous groups tasks. *Small Group Research*, 31(2), 225-244.

- Bonner, B. L., & Baumann, M. R. (2008). Informational intragroup influence: The effects of time pressure and group size. *European Journal of Social Psychology, 38*, 46-66.
- Bonner, B. L., Baumann, M. R., Lehn, A. K., Pierce, D. M., & Wheeler, E. C. (2006). Modeling collective choice: Decision-making on complex intellectual tasks. *European Journal of Social Psychology, 36*, 617-633.
- Bonner, B. L., Gonzalez, C. M. & Sommer, D. (2004). Centrality and accuracy in group quantity estimations. *Group Dynamics: Theory, Research and Practice, 8*(3), 155-165.
- Bonner, B. L. & Sillito, S. D. (2011). Leveraging member knowledge in groups: Expertise, extroversion, and feedback. *Group Dynamics: Theory, Research, and Practice, 15*(3), 233-245.
- Bonner, B. L., Sillito, S. D., & Baumann, M. R. (2007). Collective estimation: Accuracy, expertise, and extroversion as sources of intra-group influence. *Organizational Behavior and Human Decision Processes, 103*, 121-133.
- Boos, M., Scharpf, U., & Fisch, R. (1991). Eine Methode zur Analyse von Interaktionsprozessen beim Problemlösen und Entscheiden in Sitzungen. *Zeitschrift für Arbeits- und Organisationspsychologie, 35*(3), 115-121.
- Borkenau, P., & Ostendorf, F. (1993). NEO-Fünf-Faktoren Inventar (NEO-FFI) nach Costa und McCrae: Handanweisung. Göttingen: Hogrefe.
- Bortz, J. (1999). *Statistik für Sozialwissenschaftler*. (5. überarb. Aufl.). Berlin: Springer.
- Bortz, J., & Döring, N. (2006). *Forschungsmethoden und Evaluation* (4. überarb. Aufl.). Berlin: Springer.
- Brehmer, B. (1992). Dynamic decision making: Human control of complex systems. *Acta Psychologica, 81*, 211-241.
- Brehmer, B. (2005). Micro-worlds and the circular relation between people and their environment. *Theoretical Issues in Ergonomics Science, 6*(1), 73-93.
- Broadbent, D. E. (1977). Levels, hierarchies, and the locus of control. *Quarterly Journal of Experimental Psychology, 29*, 181-201.

- Broadbent, D. E. & Ashton, B. (1978). Human control of a simulated economic system. *Ergonomics*, 21(12), 1035-1043.
- Broadbent, D. E., FitzGerald, P., & Broadbent, H. P. (1986). Implicit and explicit knowledge in the control of complex systems. *British Journal of Psychology*, 77, 33-50.
- Brodbeck, F. C., & Greitemeyer, T. (2000a). A dynamic model of group performance: Considering the group members' capacity to learn. *Group Processes and Intergroup Relations*, 3, 159-182.
- Brodbeck, F. C., & Greitemeyer, T. (2000b). Effects of individual versus mixed individual and group experience in rule induction on group member learning and group performance. *Journal of Experimental Social Psychology*, 36, 621-648.
- Burke, C. S., Stagl, K. C., Salas, E., Pierce, L., & Kendall, D. (2006). Understanding team adaptation: A conceptual analysis and model. *Journal of Applied Psychology*, 91(6), 1189-1207.
- Cohen, J. (1988): *Statistical Power Analysis for the Behavioral Sciences* (2. Aufl.). Hillsdale: Lawrence Erlbaum Associates.
- Collins, B. E., & Guetzkow, H. (1964). *A social psychology of group processes for decision-making*. New York: Wiley.
- Comprehensive Meta-Analysis (Version 2.0). [Computer software]. Englewood, NJ: Biostat.
- Cosier, R. A. (1978). The effects of three potential aids for making strategic decisions on prediction accuracy. *Organizational Behavior and Human Performance*, 22, 295-306.
- Cosier, R. A. & Rechner, P. L. (1985). Inquiry method effects on performance in a simulated business environment. *Organizational Behavior and Human Decision Processes*, 36, 79-95.
- Cruz, M. G., Henningsen, D. D., & Smith, B. A. (1999). The impact of directive leadership on group information sampling, decisions, and perceptions of the leader. *Communication Research*, 26(3), 349-369.
- D'Azzo, J. J. & Houpis, C. H. (1982). *Linear control system analysis and design: Conventional and modern*. New York, NY: McGraw-Hill.

- Deci, E. L. & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York, NY: Plenum Press.
- Dentler, P. (1977). *Zum Problemlöseverhalten von Gruppen in komplexen Problembereichen*. Kiel: Unveröffentlichte Dissertation.
- Diehl, E. W. (1992). Participatory simulation software for managers: The design philosophy behind micro world creator. *European Journal of Operations Research*, 59, 210-215.
- Diehl, E. & Serman, J. D. (1995). Effects of feedback complexity on dynamic decision making. *Organizational Behavior and Human Decision Processes*, 62(2) 198-215.
- Dörner, D, Kreuzig, H. W., Reiter, F., & Stäudel, T. (1983). *Lohhausen. Vom Umgang mit Unbestimmtheit und Komplexität*. Bern: Huber.
- Dörner, D. & Reither, R. (1978). Über das Problemlösen in sehr komplexen Realitätsbereichen. *Zeitschrift für experimentelle und angewandte Psychologie* 25(4), 527-551.
- Dörner, D. & Wearing, A. J. (1995). Complex problem solving: Toward a (computer simulated) theory. In P. A. Frensch & J. Funke (Eds.), *Complex problem solving: The European perspective* (pp. 65-99). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Dörner, D. (1992). *Die Logik des Mißlingens. Strategisches Denken in komplexen Situationen*. Reinbek: Rowohlt.
- Drewes, S., Schultze, T., & Schulz-Hardt, S. (2011). Leistung in Gruppen. In D. Frey & H.-W. Bierhoff (Hrsg.) *Sozialpsychologie. Interaktion und Gruppe*. Göttingen: Hogrefe.
- Endres, J. & Putz-Osterloh, W. (1994). Komplexes Problemlösen in Kleingruppen: Effekte des Vorwissens, der Gruppenstruktur und der Gruppeninteraktion. *Zeitschrift für Sozialpsychologie*, 25(1), 54-70.
- Funke, J. (1985). Steuerung dynamischer Systeme durch Aufbau und Anwendung subjektiver Kausalmodelle. *Zeitschrift für Psychologie*, 193, 443-465.
- Funke, J. (1992). Wissen über dynamische Systeme: Erwerb, Repräsentation und Anwendung. In D. Albert, K. Pawlik, K.-H. Stapf, & W. Stroebe (Eds.), *Lehr- und Forschungstexte Psychologie Vol. 43*. Berlin: Springer.

- Funke, J. (2001a). Dynamic systems as tools for analysing human judgement. *Thinking and Reasoning*, 7(1), 69-89.
- Funke, J. (2001b). Neue Verfahren zur Erfassung intelligenten Umgangs mit komplexen und dynamischen Anforderungen. In E. Stern und J. Guthke (Hrsg.), *Perspektiven der Intelligenzforschung* (S. 89-109). Lengerich: Pabst Science Publishers.
- Funke, J. & Müller, H. (1988). Eingreifen und Prognostizieren als Determinanten von Systemidentifikation und Systemsteuerung. *Sprache & Kognition*, 7, 27-37.
- Gebauer, G. F. & Mackintosh, N. J. (2007). Psychometric intelligence dissociates implicit and explicit learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 34–54.
- Gonzalez, C., Vanyukov, P., & Martin, M. K. (2005). The use of microworlds to study dynamic decision making. *Computers in Human Behavior*, 21, 273-286.
- Hackman, J. R. & Morris, C. G. (1975). Group tasks, group interaction process, and group performance effectiveness: A review and proposed integration, In L. Berkowitz (Ed.) *Advances in experimental social psychology*, Vol.8 (pp. 47-99), New York, NY: Academic Press.
- Heineken, E., Arnold, H. - J., Kopp, A. & Soltysiak, R. (1992). Strategien des Denkens bei der Regelung eines einfachen dynamischen Systems unter verschiedenen Totzeitbedingungen. *Sprache & Kognition*, 11, 136-148.
- Henry, R. A. (1993). Group judgment accuracy: Reliability and validity of postdiscussion confidence judgments. *Organizational Behavior and Human Decision Processes*, 56, 11-27.
- Henry, R. A. Strickland, O. J., Yorges, S. L., & Ladd, D. (1996). Helping groups determine their most accurate member: The role of outcome feedback. *Journal of Applied Social Psychology*, 26, 1153-1170.
- Hofinger, G., Rek, U., & Strohschneider, S. (2006). Menschengemachte Umweltkatastrophen – Psychologische Hintergründe am Beispiel von Tschernobyl. *Umweltpsychologie*, 10(1), 26-45.

- Hollenbeck, J. R., Ilgen, D. R., Segoe, D. J., Hedlund, J., Major, D. A., & Philips, J. (1995). Multilevel theory of team decision making: Decision performance in teams incorporating distributed expertise. *Journal of Applied Psychology, 80*, 292-316.
- Howie, E., Sy, S., Ford, L., & Vicente, K. J. (2000). Human-computer interface design can reduce misperceptions of feedback. *System Dynamics Review, 16*(3), 151-171.
- Hsiao, N. (2000). Exploration of outcome feedback for dynamic decision making (Doctoral dissertation, University at Albany, State University of New York, 2000). Retrieved April 4, 2009, from <http://cc.shu.edu.tw/~nhsiao/beergame/>
- Hsiao, N. & Richardson, G. P. (1999). In search of theories of dynamic decision making: A literature review, in R. Y. Cavana et al. (Eds.): *Systems Thinking for the Next Millenium - Proceedings of the 17th International Conference of the System Dynamics Society*, C-ROM.
- Ingham, A. G., Levinger, G., Graves, J. & Peckham, V. (1974). The Ringelmann effect: Studies of group size and group performance. *Journal of Experimental Social Psychology, 10*, 371-384.
- Janis, I. L. (1982). *Psychological studies of foreign-policy decisions and fiascoes* (2<sup>nd</sup> edition). Boston: Houghton Mifflin.
- Johnson, D. W., Johnson, R. T., & Stanne, M. B., (2000). Cooperative learning methods: A meta-analysis. Minneapolis, MN: University of Minnesota. Retrieved May 19, 2010, from <http://www.tablelearning.com/uploads/File/>
- Kampmann, C. & Sterman, J. D. (1998). *Feedback complexity, bounded rationality, and market dynamics*. Working paper. System Dynamics Group. MIT Sloan School of Management, Cambridge. Retrieved from: <http://jsterman.scripts.mit.edu/>
- Kanfer, R. & Ackerman, P. L. (1989). Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition. *Journal of Applied Psychology, 74*, 657-690.
- Kerr, N. L. & Bruun, S. E. (1983). Dispensability of member effort and group motivation losses: Free-rider effects. *Journal of Personality and Social Psychology, 44*(1), 78-94.
- Kersting, M. (1999). *Diagnostik und Personalauswahl mit computergestützten Problemlöseszenarien? Zur Kriteriumsvalidität von Problemlöseszenarien und Intelligenztestes*. Göttingen: Hogrefe.



- Kickul, J., & Neuman, G. (2000). Emergent leadership behaviors: The function of personality and cognitive ability in determining teamwork performance and KSAS. *Journal of Business and Psychology, 15*(1), 27-51.
- Kieras D. E. & Bovair, S. (1984). The role of a mental model in learning to operate a device. *Cognitive Science, 8*, 255-273.
- Kleiter, G. D. (1970). Trend-control in a dynamic decision-making task. *Acta Psychologica, 34*(2-3), 387-397.
- Kluge, A. (2004). *Wissenserwerb für das Steuern komplexer Systeme*. Lengerich: Pabst Science Publishers.
- Kluwe, R. H. (1995). Single case studies and models of complex problem solving. In P. A. Frensch & J. Funke (Eds.), *Complex problem solving: The European perspective* (pp. 269-291). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Köller, O., Dauenheimer, D. G., & Strauß, B. (1993). Unterschiede zwischen Einzelpersonen und Dyaden beim Lösen komplexer Probleme in Abhängigkeit von der Ausgangsfähigkeit. *Zeitschrift für Experimentelle und Angewandte Psychologie, 40*(2), 194-221.
- Krassuski, L. (2001, 1. April). Wahre Zahl der Tschernobyl-Opfer bleibt im Dunkeln. Die Welt. Retrieved February 8, 2013 from <http://www.welt.de/wissenschaft/article13152758/>
- Kröner, S. (2001). Intelligenzdiagnostik per Computersimulation. Reihe Pädagogische Psychologie und Entwicklungspsychologie, Band 24. Münster: Waxmann.
- Kröner, S. & Leutner, D. (2002). MultiFlux - Pilotstudie für die Entwicklung eines Verfahrens zur simulationsbasierten Intelligenzdiagnostik. *Zeitschrift für Arbeits- und Organisationspsychologie, 46*, 84-88
- Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159-174.
- Larson, J. R. Jr. (2010). *In search of synergy in small group performance*. New York, NY: Psychology Press.

- Laughlin, P. R. (1980). Social combination processes of cooperative problem-solving groups on verbal intellectual tasks. In M. Fishbein (Ed.), *Progress in social psychology* (pp. 127-155). Hillsdale, NJ: Erlbaum.
- Laughlin, P. R., Bonner, B. L., & Miner, A. G. (2002). Groups perform better than the best individuals on letters-to-numbers problems. *Organizational Behavior and Human Decision Processes*, *88*, 605-620.
- Laughlin, P. R., Bonner, B. L., Miner, A. G., & Carnevale, P. J. (1999). Frames of reference in quantity estimations by groups and individuals. *Organizational Behavior and Human Decision Processes*, *80*(2), 103-117.
- Laughlin, P. R., Carey, H. R., & Kerr, N. L. (2008). Group-to-individual problem-solving transfer. *Group Processes & Intergroup Relations*, *11*(3), 319-330.
- Laughlin, P. R., & Ellis, A. L. (1986). Demonstrability and social combination processes on mathematical intellectual tasks. *Journal of Experimental Social Psychology*, *22*, 177-189.
- Laughlin, P. R., Hatch, E., C., Silver, J. S., & Boh, L. (2006). Groups perform better than the best individuals on letters-to-numbers problems: Effects of group size. *Journal of Personality and Social Psychology*, *90*(4), 644-651.
- Laughlin, P. R., VanderStoep, S. W., & Hollingshead, A. B. (1991). Collective versus individual induction: Recognition of truth, rejection of error, and collective information processing. *Journal of Personality and Social Psychology*, *61*(1), 50-67.
- Laughlin, P. R., Zander, M. L., Knieval, E. M., & Tan, T. K. (2003). Groups perform better than the best individuals on letters-to-numbers problems: Informative equations and effective strategies. *Journal of Personality and Social Psychology*, *85*(4), 684-694.
- Lee, H. L., Padmanabhan, V., & Whang, S. (1997). The bullwhip effect in supply chains. *Sloan Management Review*, *38*(3), 93-102.
- Lipshitz, R., Klein, G., Orasanu, J., & Salas, E. (2001). Taking stock of naturalistic decision making. *Journal of Behavioral Decision Making*, *14*, 332-351.
- Littlepage, G. E., Schmidt, G. W., Whisler, E. W., & Frost, A. G. (1995). An input-process-output analysis of influence and performance in problem-solving groups. *Journal of Personality and Social Psychology*, *69*(5), 877-889.

- Lorge, I., Fox, D., Davitz, J., & Brenner, M. (1958). A survey of studies contrasting the quality of group performance and individual performance, 1920-1957. *Psychological Bulletin*, 55(6), 337-372.
- Maier, F. H. & Größler, A. (2000). What are we talking about? A taxonomy of computer simulations to support learning. *System Dynamics Review*, 16(2), 135-148.
- Maier, N. R. F. (1931). Reasoning in humans. II. The solution of a problem and its appearance in consciousness. *Journal of Comparative Psychology*, 12(2), 181-194.
- Mackinnon, A. J., & Wearing, A. J. (1985). Systems analysis and dynamic decision making. *Acta Psychologica*, 58(2), 159-172.
- McClelland, D. C. (1964). *The roots of consciousness*. New York, NY: Van Nostrand.
- McCrae, R. R., & Costa, P. T., Jr. (1992). Reinterpreting the Myers-Briggs Type Indicator from the perspective of the five-factor model of personality. *Journal of Personality*, 57(1), 18-40.
- Meadows, D. (2007). A brief and incomplete history of operational gaming in system dynamics. *System Dynamics Review*, 23(2-3), 199-203.
- Meyer, B. & Scholl, W. (2009). Complex problem solving after unstructured discussion: Effects of information distribution and experience. *Group Processes & Intergroup Relations*, 12, 495-515.
- Moreland, R. L. & Myaskovsky, L. (2000). Exploring the performance benefits of group training: Transactive memory or improved communication? *Organizational Behavior and Human Decision Processes*, 82, 117-133.
- Omodei, M. M. & Wearing, A. J. (1993). Fire chief user manual. Melbourne: Department of Psychology, University of Melbourne.
- Omodei, M. M. & Wearing, A. J. (1995b). The Fire chief microworld generating program: An illustration of computer-simulated microworlds as an experimental paradigm for studying complex decision making behavior. *Behavior Research Methods, Instruments and Computers*, 27(3), 303-316.
- Osman, M. (2010). Controlling uncertainty: A review of human behavior in complex dynamic environments. *Psychological Bulletin*, 136(1), 65-86.

- Paich, M., & Sterman, J. D. (1993). Boom, bust and failures to learn in experimental markets. *Management Science*, 39(12), 1439-1458.
- Petzing, F. (1993). Untersuchung des Wissenserwerbs bei einem Unternehmensplanspiel mit Verfahren des Operations-Research. *Schriftenreihe Handeln und Entscheiden in komplexen ökonomischen Situationen: Vol. 8*. Heidelberg: Physica-Verlag.
- Preußler, W. (1996). Zur Rolle expliziten und impliziten Wissens bei der Steuerung dynamischer Systeme. *Zeitschrift für Experimentelle Psychologie*, 3, 399-434.
- Preußler, W. (1997). Effekte des Kontextes auf den Wissenserwerb bei der Steuerung eines dynamischen Systems. *Sprache & Kognition*, 16, 48-59.
- Putz-Osterloh, W. (1991). Stärken und Schwächen der Entscheidungsbildung in Gruppen im Vergleich zu Einzelpersonen. Manuskript zum Vortrag auf der 33. TeaP in Gießen 1991.
- Putz-Osterloh, W. (1993). Strategies for knowledge acquisition and transfer of knowledge in dynamic tasks. In G. Strube & K. F. Wender (Eds.), *The Cognitive Psychology of Knowledge* (p. 331-350). Amsterdam: Elsevier Science Pub.
- Putz-Osterloh, W., Bott, B., & Houben, I. (1988). Beeinflusst Wissen über ein realitätsnahes System dessen Steuerung? *Sprache & Kognition*, 7, 12-24.
- R (Version 2.11.0). Retrieved from <http://cran.r-project.org/bin/windows/base/>
- Reichert, U. & Dörner, D. (1988). Heuristiken beim Umgang mit einem "einfachen" dynamischen System. *Sprache und Kognition*, 7(1), 12-24.
- Reither, F. (1979). *Über die Selbstreflexion beim Problemlösen*. Gießen: unveröffentlichte Dissertation.
- Rigas, G. & Brehmer, B. (1999). *Mental processes in intelligence tests and dynamics decision making tasks*. London: Erlbaum.
- Rosenthal, R. (1991). Meta-analytic procedures for social research. *Applied Social Research Methods Series*, Vol. 6. London: Sage.
- Rouwette, E.A.J.A., Größler A., Vennix J.A.M. (2004). Exploring influencing factors on rationality: A literature review of dynamic decision-making studies in system dynamics. *Systems Research and Behavioral Science*, 21(4), 351-370.

- Salazar, A. J. (1995). Understanding the synergistic effects of communication in small groups: Making the most out of group member abilities. *Small Group Research*, 26(169), 169-199.
- Schaub, H. & Strohschneider, S. (1992). Die Auswirkungen unterschiedlicher Problemlöseerfahrung auf den Umgang mit einem unbekanntem komplexen Problem. *Zeitschrift für Arbeits- und Organisationspsychologie*, 36, 117-126.
- Schlemmer, A. (2009). Assembly Lab for Experiments (Versionen 2009 bis 2011) [Computer Software]. Göttingen und Hildesheim: GNU General Public License Version 3.
- Schultze, T. (2007). *Einfluss von Kompetenzfeedback auf die Gruppenleistung bei diskretionären Aufgaben* (Unveröffentlichte Diplomarbeit). Georg-August-Universität Göttingen, Göttingen.
- Schultze, T., Mojzisch, A., & Schulz-Hardt, S. (2012). Why groups perform better than individuals at quantitative judgment tasks: Group-to-individual transfer as an alternative. *Organizational Behavior and Human Decision Processes*, 118, 24-36.
- Schulz-Hardt, S., Brodbeck, F. C., Mojzisch, A., Kerschreiter, R., & Frey, D. (2006). Group decision making in hidden profile situations: Dissent as a facilitator for decision quality. *Journal of Personality and Social Psychology*, 91(6), 1080-1093.
- Schweiger, D. M. & Sandberg, W. R. (1989). The utilization of individual capabilities in group approaches to strategic decision-making. *Strategic Management Journal*, 10(1), 31-43.
- Schweiger, D. M., Sandberg, W. R., & Rechner, P. L. (1989). Experiential effects of dialectical inquiry, devil's advocacy, and consensus approaches to strategic decision making. *Academy of Management Journal*, 32(4), 745-772.
- Simon, H. A. (1982). *Models of bounded rationality*. Cambridge: MIT Press.
- Snizek, J. A., & Henry, R. A. (1989). Accuracy and confidence in group judgment. *Organizational Behavior and Human Decision Processes*, 52, 124-155.
- Snizek, J. A., & Henry, R. A. (1990). Revision, weighting, and commitment in consensus group judgment. *Organizational Behavior and Human Decision Processes*, 45, 66-84.

- Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology*, 13, 290–312.
- Srivastava, V. N. & Bhattacharya, A. S. (2007). Team work training for improving performance management: A case of Indian nuclear power plants. *IIMB Management Review*, 19(4), 459-469.
- Stasser, G., & Dietz-Uhler, B. (2001). Collective choice, judgment, and problem-solving. In M. S. Hogg & R. S. Tindale (Eds.), *Blackwell Handbook of Social Psychology* (Vol. 4, pp. 31-55). Oxford, UK: Blackwell Publishers.
- Steiner, I. D. (1972). *Group processes and productivity*. New York, NY: Academic Press.
- Sterman, J. D. (1989a). Misperceptions of feedback in dynamic decision making. *Organizational Behavior and Human Decision Processes*, 43(3), 301-335.
- Sterman, J. D. (1989b). Modeling managerial behavior: Misperceptions of feedback in a dynamic decision making experiment. *Management Science*, 35(3), 321-339.
- Sterman, J. D. (1994). Learning in and about complex systems. *System Dynamics Review*, 10(2/3), 291-330.
- Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations*. New York, NY: Doubleday.
- Süß, H. M. (1996). *Intelligenz, Wissen und Problemlösen*. Göttingen: Hogrefe.
- Taylor, H. C. & Russel, J. T. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection. *Journal of Applied Psychology*, 23, 565-578.
- Tetlock, P. E. & Levy, A. (1982). Attribution bias: On the inconclusiveness of the cognition motivation debate. *Journal of Experimental Social Psychology*, 18, 68-88.
- Vroom, V. H. & Jago, A. G. (1988). *The new leadership: Managing participation in organizations*. Englewood Cliffs, USA: Prentice-Hall.
- Wegner, D. M. (1986) Transactive memory: A contemporary analysis of the group mind. In B. Mullen & G. R. Goethals (Eds.), *Theories of group behavior* (pp. 185-208). New York, NY: Springer.

- 
- Wegner, D. M., Raymond, P., & Erber, R. (1991). Transactive memory in close relationships. *Journal of Personality and Social Psychology*, 61(6), 923-929.
- Wolfe, J. & Chacko, T. I. (1983). Team-size effects on business game performance and decision making behaviors. *Decision Sciences*, 14, 121-133.
- Zhang, Z., Hempel, P. S., Han, Y., & Tjosvold, D. (2007). Transactive memory system links work team characteristics and performance. *Journal of Applied Psychology*, 92, 1722-1730.

## Tabellenverzeichnis

Tabelle 1: <i>Schema zur Kategorisierung von Gruppenaufgaben nach Stasser und Dietz-Uhler (2001, S. 34) .....</i>	8
Tabelle 2: <i>Inhaltskategorien zur Kodierung des verbalisierbaren Systemwissens der Probanden in Experiment 1.....</i>	44
Tabelle 3: <i>Mittelwerte, Standardabweichungen und Konfidenzintervalle der mittleren absoluten prozentualen Abweichung (Mean Absolute Percent Error, MAPE) der erzielten Temperaturen von der Zieltemperatur in den Trials 2 bis 60 der Steuerungsphase 2 .....</i>	47
Tabelle 4: <i>Mittelwerte und Standardabweichungen der drei abhängigen Variablen Gesamtzahl der Annahmen, Anzahl richtiger Annahmen und Prozentsatz richtiger Annahmen für die absoluten Werte in Steuerungsphase 2 sowie für das Differenzmaß aus Phase 1 und Phase 2.....</i>	62
Tabelle 5: <i>Mittelwerte und Standardabweichungen des Prozentsatzes richtig identifizierter relativer Mitgliederkompetenz in der Realbedingung.....</i>	66
Tabelle 6: <i>Inhaltskategorien zur Kodierung des verbalisierbaren Systemwissens der Probanden in Experiment 2.....</i>	84
Tabelle 7: <i>Mittelwerte, Standardabweichungen und Konfidenzintervalle der mittleren log<sub>2</sub>-transformierten Gesamtkosten in der Steuerungsphase des Produktions- und Lagermanagementsystems in Experiment 2 .....</i>	90
Tabelle 8: <i>Mittelwerte und Standardabweichungen der drei abhängigen Variablen Gesamtzahl der Annahmen, Anzahl richtiger Annahmen und Prozentsatz richtiger Annahmen für die absoluten Werte in Steuerungsphase 2 sowie für das Differenzmaß aus Phase 1 und Phase 2 in Experiment 2 .....</i>	101



---

Tabelle 9: <i>Mittelwerte und Standardabweichungen des Prozentsatzes richtig identifizierter relativer Mitgliederkompetenz in der Realbedingung des Experimentes 2</i> .....	104
Tabelle 10: <i>Resultierende MAPE.Werte und RMSD unter den normativen Vergleichsmodellen in Experiment 1 und 3 im Vergleich</i> .....	120
Tabelle 11: <i>Differenzwerte und Ergebnisse separater Einstichproben-t-Tests gegen Null für das Differenzmaß zum individuellen Leistungszuwachs in den Steuerungsphasen 2 und 3 des Experimentes 3</i> .....	124
Tabelle 12: <i>Nach Bedingungen getrennte Mittelwerte und Standardabweichungen für die absoluten Werte der abhängigen Variablen zum Wissenserwerb in Steuerungsphase 2 sowie für das Differenzmaß aus Phase 1 und Phase 2 in Experiment 3</i> .....	130
Tabelle 13: <i>Nach Bedingungen getrennte Mittelwerte und Standardabweichungen für die absoluten Werte der abhängigen Variablen zum Wissenserwerb in Steuerungsphase 2 sowie für das Differenzmaß aus Steuerungsphase 3 und Steuerungsphase 1 in Experiment 3</i> .....	132
Tabelle 14: <i>Differenzmaßwerte und Ergebnisse separater Einstichproben-t-Tests gegen Null für den individuellen Leistungszuwachs in der Steuerungsphase 2 des Experiments 1</i> .....	195
Tabelle 15: <i>Differenzmaßwerte und Ergebnisse separater Einstichproben-t-Tests gegen Null für den individuellen Leistungszuwachs in der Steuerungsphase 2 des Experiments 2</i> .....	232

## Abbildungsverzeichnis

<i>Abbildung 1.</i> Formel zur Divergenz zwischen tatsächlicher Gruppenleistung und Gruppenpotenzial in Anlehnung an Hackman und Morris (1975). .....	21
<i>Abbildung 2.</i> Screenshot der auf dem Computerbildschirm angezeigten Ein- und Ausgabeoberfläche zur Systemsteuerung in Experiment 1. Dargestellt ist der Ausgangszustand vor Eingabe der ersten Stellradeinstellung in Steuerungsphase 1 bzw. Steuerungsphase 2.....	33
<i>Abbildung 3.</i> Dem dynamischen System in Experiment 1 zugrunde liegende Systemgleichungen und Erläuterungen der Gleichungsparameter. (konstante Ausprägung der Werte bzw. Ausgangswerte zum Zeitpunkt $i = 1$ in eckigen Klammern). .....	35
<i>Abbildung 4.</i> Blockdiagramm des dynamischen Systems in Experiment 1. ....	36
<i>Abbildung 5.</i> Verlauf der Temperaturwerte über die 60 Trials der Steuerungsphase 2 in Experiment 1, die aus der konstanten systembedingt minimalen, der systembedingt maximalen Stellradeinstellung, aus der optimal rule oder aus der no control rule resultierten.....	38
<i>Abbildung 6.</i> Schematische Darstellung des Versuchsablaufes von Experiment 1.....	42
<i>Abbildung 7.</i> Verlauf der MAPE-Werte über die 60 Trials der Steuerungsphase 2, die aus der Steuerung in den drei Bedingungen sowie aus den Benchmarks der optimal rule und der no control rule resultierten. ....	48
<i>Abbildung 8.</i> MAPE der Temperaturwerte der drei Bedingungen im Zeitverlauf von Phase 2 des Experiments 1. ....	51
<i>Abbildung 9.</i> Ausmaß der individuellen Leistungssteigerung in Experiment 1 für das gemäß der Individualleistung in Phase 1 beste, mittlere und schlechteste Gruppenmitglied. ....	58
<i>Abbildung 10.</i> Schematische Darstellung der auf dem Computerbildschirm angezeigten Ein- und Ausgabeoberfläche zur Systemsteuerung. Dargestellt ist der Ausgangszustand vor Eingabe des ersten Wertes für die Änderung der Produktionsrate. ....	77

---

<i>Abbildung 11.</i> Dem dynamischen in Experiment 2 zugrunde liegende Kostengleichung und Erläuterungen der Gleichungsparameter im Überblick (Gleichungssystem übernommen aus Diehl & Sterman, 1995). .....	80
<i>Abbildung 12.</i> Struktur der dynamischen Systemsteuerungsaufgabe in Experiment 2 (In Anlehnung an die Darstellung Fig. 2.1, S.65 in Diehl, 1992). .....	81
<i>Abbildung 13.</i> Mittlere erzielte Gesamtkosten pro Bedingung für jeden der 40 Trials der Steuerungsphase 2.....	92
<i>Abbildung 14.</i> Ausmaß der individuellen Leistungssteigerung für das gemäß der Individualleistung in Phase 1 beste, mittlere und schlechteste Gruppenmitglied (GM) in Experiment 2. G-I Transfer wurde über ein Differenzmaß aus den log2-transformierten Gesamtkosten der Phase 1 und der Phase 2 bestimmt. ....	99
<i>Abbildung 15.</i> Schematische Darstellung des Ablaufes von Experiment 3. ....	113
<i>Abbildung 16.</i> Verlauf der Temperaturkurven über die 60 Trials der Steuerungsphase 2, die aus der Steuerung in den drei Bedingungen in Experiment 3 resultierten. ....	117
<i>Abbildung 17.</i> Ausmaß der individuellen Leistungssteigerung für das gemäß der Individualleistung in Phase 1 beste, mittlere und schlechteste Gruppenmitglied (GM) in Experiment 3. G-I Transfer wurde über ein Differenzmaß aus dem MAPE der Phase 1 und dem MAPE der Phase 2 bestimmt. ....	123
<i>Abbildung 18.</i> Ausmaß des G-I Transfers für das gemäß der Individualleistung in Phase 1 beste, mittlere und schlechteste Gruppenmitglied in Experiment 3. G-I Transfer wurde über ein Differenzmaß aus dem MAPE der Phase 1 und dem MAPE der Phase 3 bestimmt. ....	127
<i>Abbildung 19.</i> Subjektives Wohlbefinden in der Explorationsphase, der Steuerungsphase 1 und der Steuerungsphase 2 getrennt nach Bedingungen in Experiment 1. ....	203
<i>Abbildung 21.</i> Subjektive Anstrengung bei der Systemexploration- bzw. steuerung in der Explorationsphase, der Steuerungsphase 1 und der Steuerungsphase 2 getrennt nach Bedingungen in Experiment 1. ....	205
<i>Abbildung 22.</i> Subjektives Wohlbefinden in der Explorationsphase, der Steuerungsphase 1 und der Steuerungsphase 2 getrennt nach Bedingungen in Experiment 2. ....	235

---

<i>Abbildung 23.</i> Subjektive Wichtigkeit, das System zu verstehen in der Explorationsphase, der Steuerungsphase 1 und der Steuerungsphase 2 getrennt nach Bedingungen in Experiment 2. ....	236
<i>Abbildung 24.</i> Subjektive Anstrengung bei der Systemexploration- bzw. steuerung in der Explorationsphase, der Steuerungsphase 1 und der Steuerungsphase 2 getrennt nach Bedingungen in Experiment 2. ....	237
<i>Abbildung 25.</i> Subjektives Wohlbefinden in der Explorationsphase, der Steuerungsphase 1 und der Steuerungsphase 2 getrennt nach Bedingungen in Experiment 3. ....	247
<i>Abbildung 26.</i> Subjektive Wichtigkeit, das System zu verstehen in der Explorationsphase, der Steuerungsphase 1 und der Steuerungsphase 2 getrennt nach Bedingungen in Experiment 3. ....	248
<i>Abbildung 27.</i> Subjektive Anstrengung in der Explorationsphase, der Steuerungsphase 1 und der Steuerungsphase 2 getrennt nach Bedingungen in Experiment 3. ....	249



## Anhang

- Anhang A      Detaillierte Ablaufbeschreibung des Experimentes 1
- Anhang B      Versuchsleitungsanweisung zu Experiment 1
- Anhang C      Kodieranweisung zu den Wissensfragen in Experiment 1 und 3
- Anhang D      Ergänzende Analysen zu Lerneffekten auf individueller Ebene in Experiment 1
- Anhang E      Explorative Analyse: Zusammenhang von Persönlichkeitsfaktoren mit der Steuerungsleistung und Gewichtung der individuellen Urteile
- Anhang F      Explorative Analyse: Zusammenhang von motivationalen Variablen und Steuerungsleistung in Experiment 1
- 
- Anhang G      Detaillierte Ablaufbeschreibung des Experimentes 2
- Anhang H      Versuchsleitungsanweisung zu Experiment 2
- Anhang I      Schriftliches Versuchsmaterial zu Experiment 2
- Anhang J      Kodieranweisung zu den Wissensfragen in Experiment 2
- Anhang K      Ergänzende Analysen zu Lerneffekten auf individueller Ebene in Experiment 2
- Anhang L      Explorative Analyse: Zusammenhang von motivationalen Variablen und Steuerungsleistung in Experiment 2
- 
- Anhang M      Versuchsleitungsanweisung zu Experiment 3
- Anhang N      Explorative Analyse: Zusammenhang von motivationalen Variablen und Steuerungsleistung in Experiment 3 und Experiment übergreifende Analysen
- Anhang O      Lebenslauf

## Detaillierte Ablaufbeschreibung des Experimentes 1

Zu einem Erhebungstermin wurden bis zu 11 Probanden eingeladen, die von zwei oder drei Versuchsleitern betreut wurden. Zu Beginn des Experimentes wurden die Probanden in einem großen Versuchsraum vor separate Laptops gesetzt und von der Versuchsleitung begrüßt und mündlich über den Ablauf des Experimentes informiert. Der einleitende mündliche Instruktionstext war standardisiert und wird im Folgenden beschrieben. Das Experiment wurde den Probanden als eine Untersuchung zum Thema „Systemsteuerungsaufgaben“ vorgestellt und als Untersuchungsziel wurde Interesse daran, wie Personen individuell oder in der Gruppe diesen bestimmten Aufgabentyp bearbeiten, genannt. Für die voraussichtliche Untersuchungsdauer wurden circa 45 bis 75 Minuten, in Abhängigkeit von der Versuchsbedingung und dem individuellem Arbeitstempo, angekündigt. Den Probanden wurde mitgeteilt, dass das gesamte Experiment am Laptop stattfände und sich in fünf Teile gliedern würde. Anschließend wurde den Probanden gesagt, dass Sie zunächst einige demographische Angaben zu machen und einige Fragen zu ihrer Person zu beantworten hätten. Die personbezogenen Fragen bestanden aus der Originalinstruktion und den vollständigen Items des NEO-Fünf-Faktoren Inventars (NEO-FFI) in der deutschen Version von Borkenau und Ostendorf (1993). Dass es sich um diesen Persönlichkeitsfragebogen handelte, wurde allerdings nicht mitgeteilt. Den Probanden wurde gesagt, dass ihnen anschließend das zu steuernde System schriftlich erläutert würde und dass sie sich in 20 Übungsdurchgängen mit der Funktionsweise vertraut machen könnten. Es ginge um die Steuerung eines Kühlsystems in einem Atomkraftwerk. Dabei handele es sich um eine Optimierungsaufgabe und das Ziel bestünde darin so schnell wie möglich einen Optimalwert zu erreichen und ihn dann möglichst gut beizubehalten. Es sollten also die Temperaturabweichungen vom Optimalwert über alle Steuerungsdurchgänge möglichst gering gehalten werden. Die an die Systemexploration anschließende Steuerungsphase 1 wurde den Probanden als individuelle Arbeitsphase angekündigt, in der sie über 60 Durchgänge das System nach der bestimmten Zielvorgabe steuern sollen. Nach dieser Phase würde die persönliche Steuerungsleistung der Probanden erfasst werden und sie würden von der Versuchsleistung in der vierten Arbeitsphase entweder einer individuellen oder einer Gruppenarbeitsbedingung zugewiesen. Anschließend wurden den Probanden die drei Bedingungen kurz erläutert, damit sie sich nicht dadurch ablenken ließen, dass der Versuchsverlauf für die zum Erhebungstermin anwesenden Personen unterschiedlich sein konnte und einige Probanden in der Realgruppenbedingung während des Versuchs den Raum verließen.

Zur Individualbedingung wurde den Probanden mitgeteilt, dass sie in dieser Arbeitsbedingung das System erneut über 60 Durchgänge alleine steuern würden. Die Nominalgruppenbedingung wurde als eine Gruppenbedingung beschrieben, in der ihr Computer über einen Server mit zwei anderen Versuchspersonencomputern verbunden sei. Hierbei würde jedes Gruppenmitglied räumlich getrennt ohne Absprache mit den anderen 60 individuelle Steuerungsurteile vornehmen. Der Server integriere die Einzeleingaben pro Durchgang zu einem mittleren Gruppenurteil und melde diesen Gruppenwert und die resultierende Systembeeinflussung nach jedem einzelnen der 60 Durchgänge zurück. Dann erfolge die Eingabe für den nächsten Durchgang. Die Realgruppenbedingung wurde als eine zweite Gruppenbedingung vorgestellt, in der die Probanden in Dreiergruppen auf die Nachbarräume aufgeteilt würden und für jeden von 60 Steuerungsdurchgängen zunächst ein individuelles Urteil treffen und dann in der Gruppe diskutieren und ein Gruppenurteil festlegen sollten. Für den eingegebenen Gruppenwert erhielten sie wie gewohnt eine Systemrückmeldung über die resultierende Temperatur. Der fünfte Teil des experimentellen Ablaufes wurde als ein allgemeiner Fragebogen benannt, den die Probanden individuell am PC ausfüllen würden.

Die Versuchsleitung informierte außerdem darüber, dass die Gruppen so zusammengestellt werden würden, dass die durchschnittliche Leistung der Gruppenmitglieder möglichst gleich stark sei. Für das Experiment gäbe es eine Grundvergütung von 5 Euro bzw. eine Versuchspersonenstunde und einen leistungsabhängigen Bonus zwischen 0 und 5 Euro, der davon abhinge, wie gut sie das System nach der Zielvorgabe in der zweiten Steuerungsphase steuern würden. Bestimmt würde die mittlere prozentuale Abweichung der 60 erreichten Temperaturwerte von der anzustrebenden Optimaltemperatur. Ergänzend erwähnte die Versuchsleitung, dass es sehr hilfreich sei, sich in der ersten individuellen Phase anzustrengen, um in der zweiten Phase gut abzuschneiden. Zum Abschluss der mündlichen Instruktion folgte der Hinweis zum selbstständigen Durchlaufen des computergestützten Experimentes. Die Probanden sollten darauf achten, dass sie mit der Maus immer ganz nach unten auf einer Fragebogen bzw. Eingabeseite scrollten und dann auf den „Weiter“-Button klickten, um zur nächsten Seitenanzeige zu gelangen. Falls der Button nicht funktionierte, hätten sie vergessen, eine Frage zu beantworten. Ergänzt wurde, dass die Probanden ein „x“ in das Antwortfeld setzen sollten, falls sie eine offene Frage nicht beantworten wollen würden. Im Anschluss an die mündliche Instruktion konnten die Probanden Verständnisfragen stellen. Dann begannen sie an ihren Laborlaptops mit der Eingabe der demographischen Angaben. Während des weiteren Verlaufes auftretende Fragen wurden durch Handzeichen der Probanden



angezeigt und individuell von der Versuchsleitung beantwortet. Fragen zur Funktionsweise des dynamischen Systems waren von der Beantwortung ausgenommen.

**Demographische Angaben und NEO-Fünf-Faktoren Inventar.** Der Ablauf des computergestützten Experimentes war für alle Probanden bis zur Zuordnung zu den Bedingungen identisch. Die Bearbeitung der Fragen und Steuerungseingaben erfolgte bis zur Einteilung der Gruppen im individuellen Tempo. Auf der Startseite lasen die Probanden: „Liebe/r VersuchsteilnehmerIn, vielen Dank, dass Sie an unserer Untersuchung zum Thema „Systemsteuerungsaufgaben“ teilnehmen. Zunächst einmal bitten wir Sie um einige demographische Angaben und Angaben zur Beschreibung Ihrer Person.“ Darunter konnten die Probanden per Mausclick ihr Geschlecht auswählen und in Eingabefeldern Angaben zu Alter, Semesterzahl und Studienfach machen. Mit einem Klick auf den „Weiter“-Button gelangten die Probanden zum nächsten Eingabefenster. Hier wurde ihnen die vollständige Instruktion des *NEO-Fünf-Faktoren Inventars* (NEO-FFI) dargeboten. Dieser Persönlichkeitsfragebogen erfasst die Dimensionen Neurotizismus, Extraversion, Offenheit, Verträglichkeit und Gewissenhaftigkeit (Borkenau & Ostendorf, 1993). Den Probanden wurde die fünfstufige Skala erläutert und anschließend wurden untereinander alle 60 Items zur Selbstbeschreibung der eigenen Person angezeigt. Per Mausclick konnten die Probanden für jedes Item auf der fünfstufigen Skala von -2 (*Starke Ablehnung*) bis 2 (*Starke Zustimmung*) den Zustimmungswert auswählen, der am ehesten auf sie zutrifft. Nach Beantwortung des Persönlichkeitsinventars wurde über den „Weiter“-Button das nächste Ansichtsfenster aufgerufen. Hier lasen die Probanden unter der Überschrift „Einleitung Systemsteuerungsaufgabe“, dass sie sich mit der Lösung von Systemsteuerungsaufgaben befassen würden. Gegenstand dieser Aufgaben sei die Steuerung eines Kühlsystems in einem Atomkraftwerk. Zunächst würden Sie die Funktionsweise des Systems in einer Reihe von Durchgängen testen können, um ein Gefühl für die Art der Steuerung zu bekommen. Anschließend würden Sie das System im Hinblick auf eine bestimmte Zielerreichung steuern. Nach dieser ersten individuellen Arbeitsphase würden Sie von der Versuchsleitung einer der experimentellen Bedingungen zugeordnet und steuerten das System in einer zweiten Arbeitsphase entweder weiterhin individuell oder in einer Dreiergruppe. Den Probanden wurde erklärt, dass sich Ihr Versuchspersonenhonorar danach richten würde, wie gut Sie das System in der zweiten Arbeitsphase steuerten. Das leistungsabhängige Honorar, zusätzlich zu ihren 5 Euro bzw. 1 Versuchspersonen-Stunde, läge zwischen 0 und 5 Euro. Bei Fragen könnten Sie sich jederzeit an den/die Versuchsleiter/in wenden. Abschließend wurde ihnen viel Spaß gewünscht.

**Explorationsphase.** Durch Klick auf den „Weiter“-Button wurde auf der nächsten Seite unter der Überschrift „Übungsphase/Instruktion“ das Kühlsystemszenario vorgestellt. In einem weiteren Absatz lasen die Probanden, dass sie im Folgenden 20 Stellradeinstellungen vornehmen könnten, um zu explorieren, wie das Kühlsystem funktioniert. Nach jeder Stellradeinstellung würde ihnen die aktuelle Temperatur im Kühlsystem angezeigt. Nach dieser ersten Explorationsphase würden zwei Durchgänge folgen, in denen sie das Kühlsystem im Hinblick auf eine bestimmte Zielsetzung steuern sollten. Falls Sie Fragen hätten, könnten Sie sich gerne jederzeit an die Versuchsleitung wenden! Wenn Sie nun auf „Weiter“ klicken würden, könnten Sie als Sicherheitsbeauftragte/r in 20 Durchgängen die Funktionsweise des Kühlsystems erproben. Darauf folgend wurde die Eingabeoberfläche präsentiert. Nach der zwanzigsten Eingabe erschien der Hinweis auf dem Monitor, dass die Testphase mit Durchgang 20 beendet sei, und dass die Probanden bitte „Weiter“ drücken sollten um fortzufahren. Die folgende Eingabeseite enthielt Fragen zu Wohlbefinden, Motivation und zum Vorgehen und Systemverständnis während der Explorationsphase. Zunächst wurden drei Fragen gestellt, die durch Mausklick auf einer jeweils siebenstufigen Skala zu beantworten waren: Erstens wurden die Probanden gefragt, wie wohl sie sich während der Abgabe der letzten 20 Systemsteuerungsurteile gefühlt hätten (Einstufung von 1 = *sehr unwohl* bis 7 = *sehr wohl*). Zweitens sollten die Probanden einschätzen, wie wichtig es Ihnen sei, das System der Temperatursteuerung zu verstehen (Einstufung von 1 = *sehr unwichtig* bis 7 = *sehr wichtig*). Drittens sollten sie angeben wie viel Mühe sie sich während der Abgabe der 20 Steuerungsurteile während der Systemexploration gegeben hatten (Einstufung von 1 = *sehr wenig Mühe* bis 7 = *sehr viel Mühe*). Anschließend sollten die Probanden ihre Explorationsstrategien in einem Textfeld beschreiben. Hierzu wurden sie gefragt, wie sie in der ersten Explorationsphase des Systems mit den 20 Zeittakten vorgegangen seien. Außerdem wurde die Revidierung von System bezogenen Hypothesen über die Frage: „Haben Sie Ihre Annahme über die Funktionsweise des Kühlsystems während der Systemexploration mit dem Stellrad geändert?“ mit den Antwortalternativen *Ja, mehrmals*, *Ja, einmal* oder *Nein* erfasst. Abschließend erfolgte die Frage nach den konkreten Annahmen über den Zusammenhang von Stellradeinstellung und Kühltemperatur. Hierzu fanden die Probanden drei Textfelder für offene Antworten vor, die mit *Annahme 1* bzw. *Annahme 2* und *Annahme 3* bezeichnet waren. Die Probanden wurden bei vorliegender Änderung ihrer Annahmen im Verlauf der Steuerung gebeten, sie zeitlich so zu ordnen, dass die aktuell gültige Annahme an letzter Stelle stünde. Wenn Sie weniger als 3 Annahmen hätten, sollten Sie ein "x" in die nicht benötigten Textfelder eintragen.

**Steuerungsphase 1.** Mit Klick auf den „Weiter“-Button erschien die folgende Instruktion für die erste individuelle Steuerungsphase: „Im Atomkraftwerk kommt es zu einem Kurzschluss. Das moderne Kühlsystem wird nicht wie vorgesehen automatisch durch Notstromaggregate am Laufen gehalten. Dieser Fehler kann zu einer Kernschmelze führen und stellt eine ernsthafte Bedrohung dar! Die Behebung des Defektes durch Elektriker wird einige Zeit in Anspruch nehmen. Bis dahin ist es Ihre Aufgabe den Reaktorraum kühl zu halten, was Sie mit Hilfe des älteren, unabhängigen Kühlsystems tun können. Sie sollen das Stellrad über 60 Zeittakte so einstellen, dass möglichst schnell und dauerhaft eine Temperatur von +4°C in der Kühlleitung herrscht. Zu Beginn Ihrer manuellen Steuerung liegt die Temperatur im Kühlsystem bereits bei ca. +14°C. Nach jeder Stellradeinstellung wird Ihnen die aktuelle Temperatur im Kühlsystem angezeigt. Verändern Sie die Stellradeinstellung so, dass die optimale Temperatur von 4°C erreicht bzw. beibehalten wird. Es sind Werte von 0 bis 200 einstellbar. Klicken Sie nach jedem eingegebenen Stellradwert auf ‚OK‘ (die neue Temperatur wird ausgegeben) und anschließend auf ‚zur nächsten Einstellung‘.“

Die Probanden gaben dann 60 Steuerungsurteile mit dem Ziel ein, möglichst schnell und dauerhaft eine Temperatur von 4°C zu erreichen. Im Anschluss an die Steuerungsphase wurden die nach der Explorationsphase gestellten Fragen zu Motivation und Annahmen über Systemzusammenhänge sinngemäß erneut gestellt, wobei „Systemexploration“ durch „Systemsteuerung“ ersetzt wurde und von 60 statt 20 Durchgängen die Rede war.

**Zuteilung zu experimenteller Bedingung.** Nach einem Klick auf den „Weiter-Button“ erschien der Hinweis „Bitte rufen Sie jetzt die Versuchsleitung!“ Alle Versuchspersonenlaptops waren über Wireless LAN mit einem Serverlaptop vernetzt. Das Programm ALEX war so programmiert, dass auf diesem Laptop der Versuchsleistung tabellarisch der Fortschritt aller Probanden im Experimentverlauf inklusive der konkreten Steuerungsurteile und Temperatursausgaben angezeigt wurde. Darüber hinaus wurde automatisch der individuelle MAPE der Probanden berechnet. Die Versuchsleitung ordnete jeweils drei Probanden leistungsheterogenen Gruppen zu. Das Kriterium der Leistungsheterogenität war erfüllt, wenn die individuellen MAPE-Werte der drei Gruppenmitglieder bei allen paarweisen Vergleichen mindestens eine Differenz von .10 aufwiesen. Zusammengestellte Gruppen wurden zufällig der Nominal- oder Realbedingung zugewiesen. Probanden, die aufgrund der Probandenzahl bzw. ihres MAPE-Wertes keiner Gruppe zugeordnet werden konnten, wurden der Individualbedingung zugewiesen. Die

Versuchsleiter trugen durch passwortgeschützte Eingabe an den Probandenlaptops die Bedingungszuordnung ein und gaben gegebenenfalls die Gruppenidentifikationsnummer an. Diese Nummer ermöglichte dem Programm in der Nominalbedingung die individuellen Eingabewerte über den Serverrechner zu einem Gruppenmittelwert zu verrechnen und den drei jeweiligen Probandenlaptops zurückzumelden.

**Steuerungsphase 2 und abschließende Fragen in Einzelarbeit in der Individualbedingung.** Die Instruktion zur zweiten Steuerungsphase für die Individualbedingung war bis auf die Ergänzung, dass ein „erneuter Zwischenfall im Kraftwerk“ aufgetreten sei, identisch zur Instruktion für die erste Steuerungsphase. Alle Probanden in der Individualbedingung trafen erneut 60 Steuerungsurteile und beantworteten im Anschluss den Fragebogen zu ihrer Motivation und ihren Annahmen über die Zusammenhänge zwischen den Systemvariablen. In einem Abschlussfragebogen wurden in der Individualbedingung drei Fragen gestellt: Die Probanden wurden erstens gefragt, wie sie bei der Bearbeitung vorgegangen seien und ob sie zum Beispiel spezifische Erfahrungen aus der ersten individuellen Arbeitsphase genutzt hätten. Hierzu konnten Sie unbegrenzt Eingaben in einem Textfeld vornehmen. Zweitens wurde die subjektive Steuerungskompetenz erfragt, wobei die Probanden sich auf einer Skala von „1 = sehr geringe Kompetenz“ bis „7 = sehr hohe Kompetenz“ einschätzen sollten. Drittens sollten die Probanden in einem letzten Textfeld das angenommene Ziel dieser Untersuchung benennen.

**Steuerungsphase 2 und abschließende Fragen in Einzelarbeit in der Nominalbedingung.** Die Instruktion in der Nominalbedingung unterschied sich von der individuellen Steuerungsanweisung. Den Probanden in der Nominalbedingung wurde angezeigt, dass die Kraftwerksleitung beschlossen hätte, dass ab sofort ein Sicherheitsteam, das aus ihnen und zwei weiteren Sicherheitsbeauftragten bestünde, bei Zwischenfällen eingreifen sollte. Sie saßen an getrennten Steuerungsmonitoren und würden alle drei unabhängig voneinander Urteile über die vorzunehmenden Stellradeinstellungen abgeben. Ein erneuter Zwischenfall im Kraftwerk erfordere Ihren Einsatz als Sicherheitsbeauftragte: Das moderne Kühlsystem wäre ausgefallen. Um dauerhaft die optimale Kühltemperatur von 4°C in der Kühlleitung herzustellen, sei es erforderlich, dass Sie erneut die manuelle Steuerung des Ersatzkühlsystems übernähmen. Für Ihre Steuerungsentscheidungen in der Gruppe sei vereinbart worden, dass jeder individuell für einen Zeittakt auf seinem Monitor festhielte, welche Stellradeinstellung er vornehmen würde. Hätten alle Ihr individuelles Urteil getroffen, würden die drei Werte vom

Computersystem zu einem Wert zusammengefasst. Sie erhielten eine Systemrückmeldung über diesen endgültigen Stellradwert und dann wie gewohnt eine Rückmeldung über die resultierende aktuelle Temperatur im Kühlsystem. Nach erfolgter Rückmeldung würde für den folgenden Zeittakt individuell von Ihnen und den anderen zwei Sicherheitsbeauftragten ein Urteil über die nächste Stellradeinstellung getroffen. Aus Erfahrung wüssten Sie, dass es 60 Zeittakte dauern würde, bis das automatische System wieder einsatzbereit sei. Sie könnten nun mit der ersten Stellradeinstellung beginnen. Sobald alle drei Gruppenmitglieder ihre Stellradwerte für den Zeittakt eingegeben hätten, würde Ihnen der integrierte Stellradwert und die resultierende aktuelle Temperatur im Kühlsystem angezeigt.

Die Steuerungsoberfläche unterschied sich in zwei Punkten von der individuellen Steuerungsansicht. Nach Eingabe des individuellen Steuerungsdurchgangs wurde neben dem Eingabefeld auf dem Probandenlaptop so lange „Bitte warten“ angezeigt, bis alle drei Mitglieder der entsprechenden Nominalgruppe ihr Steuerungsurteil eingegeben hatten. Dann erfolgte die Rückmeldung der aus dem Mittelwert resultierenden Temperatur mit der Erläuterung „Der letzte integrierte Stellwert aus den drei individuellen Urteilen Ihrer Gruppe lautet:“.

Die Probanden in der Nominalbedingung beantworteten nach der zweiten Steuerungsphase denselben Fragenkomplex zu Motivation und Systemannahmen wie die Probanden in der Individualbedingung. In den abschließenden Fragen wurden sie außer nach dem Untersuchungsziel noch gefragt, wie wohl sie sich in der Gruppe gefühlt hätten und wie wichtig es Ihnen gewesen wäre, dass die Einzelbeiträge aller drei Gruppenmitglieder in gleichem Maße in das Gruppenurteil einfließen. Mit den Angaben auf einer Skala von „1 = sehr unwohl bzw. unwichtig“ bis „7 = sehr wohl bzw. sehr wichtig“ zu den letzten beiden Fragen sollte eine Überprüfung der Akzeptanz für die Nominalgruppensituation und die automatisierte Mittelung der Beiträge, sowie ein Vergleich zum Wohlbefinden in der Realbedingung ermöglicht werden.

**Steuerungsphase 2 und abschließende Fragen in Einzelarbeit in der Realbedingung.** Die Probanden, die von der Versuchsleitung der Realbedingung zugeordnet worden waren, wechselten unter Anleitung eines Versuchsleiters<sup>92</sup> mit ihren Laptops in einen separaten Versuchsraum. Hier setzten sie sich um einen

---

<sup>92</sup> Die Versuchsleistung hatten sowohl weibliche als auch männliche studentische Hilfskräfte inne. Aus Gründen der Lesbarkeit wird im Folgenden konsistent und ausschließlich die männliche Form verwendet. Diese Formulierung schließt die Versuchsleiterinnen ein.

Gruppenarbeitstisch. Der Versuchsleiter wies den Gruppenmitgliedern die Nummern 1 bis 3 zu, indem Pappkarten mit der entsprechenden Nummer vor die Laptops der Probanden gestellt wurden. Den Probanden wurde erklärt, dass die Nummern entsprechend der Probandendurchlaufnummern im Experiment zugeteilt würden. Beispielsweise würde der Proband mit der Durchlaufnummer 120 die 1, der Proband mit der 123 die 2 und der Proband mit der 126 die 3 erhalten. Die Probanden wurden gebeten sich die Zuordnung von Nummern und Personen zur Beantwortung später folgender Fragen zu merken. Anschließend lasen die Probanden die Instruktion der zweiten Systemsteuerungsphase für die Realbedingung.

Demnach hätte die Kraftwerkleitung beschlossen, dass ab sofort ein Sicherheitsteam, das aus Ihnen und zwei weiteren Sicherheitsbeauftragten bestehe, bei Zwischenfällen eingreift. Sie säßen zu dritt in einem Raum, wobei jeder von Ihnen einen Steuerungsmonitor vor sich habe. Ein erneuter Zwischenfall im Kraftwerk erfordere Ihren Einsatz als Sicherheitsbeauftragte: Das moderne Kühlsystem sei ausgefallen. Um dauerhaft die optimale Kühltemperatur von 4°C in der Kühlleitung herzustellen, sei es erforderlich, dass sie erneut die manuelle Steuerung des Ersatzkühlsystems übernehmen würden. Für ihre Steuerungsentscheidungen in der Gruppe sei vereinbart worden, dass zunächst jeder individuell für einen Zeittakt auf seinem Monitor festhalten sollte, welche Stellradeinstellung er vornehmen würde. Hätten alle Ihr individuelles Urteil getroffen, sollte in der Gruppe über eine gemeinsame Stellradeinstellung diskutiert werden. Diesen endgültigen Stellradwert sollten sie dann alle auf ihren Laptops eingeben. Sie würden dann wie gewohnt eine Systemrückmeldung über die resultierende aktuelle Temperatur im Kühlsystem erhalten. Nach erfolgter Rückmeldung würde dann für den nächsten Zeittakt wieder erst individuell und dann in der Gruppe eine Entscheidung über eine Stellradeinstellung getroffen werden. Aus Erfahrung wüssten Sie, dass es 60 Zeittakte dauern würde, bis das automatische System wieder einsatzbereit sei. Sie könnten nun mit der ersten Stellradeinstellung (erst die individuell geplante aller drei Gruppenmitglieder, dann die von der Gesamtgruppe entschiedene) beginnen. Nach jeder in der Gruppe beschlossenen Stellradeinstellung würde Ihnen die aktuelle Temperatur im Kühlsystem angezeigt.

Die Versuchsleitung betonte anschließend, dass es wichtig sei, sich vor und während der individuellen Urteilsbildung nicht auszutauschen oder auf den Bildschirm der anderen Gruppenmitglieder zu schauen. Wenn alle ihr individuelles Urteil eingegeben hätten, sollten sie Ihre Vorschläge in der Form: „Ich habe bei Durchgang x den Wert y

eingegeben“ austauschen. Außerdem könnten sie ihre Ideen über die Funktionsweise des Systems und die optimale Eingabe diskutieren und sollten sich auf einen Gruppenwert einigen, den Sie alle drei in ihre Laptops eingeben sollten. Der individuelle Wert im Eingabefeld verschwände dann, wenn Sie auf „OK“ klickten. Dann könnten Sie den Gruppenwert eingeben und erhielten wie üblich eine Temperaturrückmeldung auf Basis der eingegebenen Gruppenurteile. Die Probanden sollten außerdem durch regelmäßiges Nennen der Durchgangsnummer sicherstellen, dass sie sich alle beim selben der 60 Durchgänge befänden. Falls sich einer in Ihrer Gruppe bei der Eingabe des Gruppenurteils vertippen würde, sollte er sich für die folgenden Durchgänge an den resultierenden Temperaturen der beiden anderen Gruppenmitglieder orientieren. Einmal getätigte Eingaben könnten nicht mehr rückgängig gemacht werden.<sup>93</sup>

Ein Versuchsleiter blieb während der Gruppenphase im Raum, um die Einhaltung des vorgeschriebenen Vorgehens zu gewährleisten. Eine Zeitvorgabe für die Gruppenarbeitsphase bestand nicht, allerdings empfanden die Probanden leichten Zeitdruck, da ihnen anfangs eine ungefähre Gesamtdauer des Experimentes von 75 Minuten angekündigt worden war. Nach alternierender Abgabe der 60 Individual- und Gruppenurteile beantworteten die Probanden individuell, ohne untereinander zu kommunizieren, die Fragen zu Motivation und Annahmen über die Systemzusammenhänge.

In den abschließenden Fragen sollten die Probanden in der Realbedingung, genauso wie in der Nominalbedingung, auf einer Skala von 1 bis 7 einschätzen, wie wohl sie sich in der Gruppe gefühlt hätten (von 1 = *sehr unwohl* bis 7 = *sehr wohl*) und wie wichtig es Ihnen gewesen wäre, dass die Einzelbeiträge aller drei Gruppenmitglieder in gleichem Maße in das Gruppenurteil einfließen (von 1 = *sehr unwichtig* bis 7 = *sehr wichtig*). Zwei zusätzliche Fragen ergänzten den Fragenblock zur allgemeinen Zusammenarbeit in der Realgruppe. Die Probanden sollten zum einen auf einer Skala von 1 bis 7 einschätzen, inwiefern ihre Beiträge von den anderen Gruppenmitgliedern angemessen berücksichtigt worden seien (von 1 = *gar nicht* bis 7 = *h*). Zum anderen wurden die Probanden gefragt, wie sie bei der Gruppenarbeit vorgegangen

---

<sup>93</sup> Fehlerhafte Eingaben von Gruppenurteilen, die daran erkennbar waren, dass ein Urteil nicht mit den zwei anderen übereinstimmte, wurden vor der Datenauswertung bereinigt. Dieses Vorgehen erschien sinnvoller, als den Probanden eine Korrekturfunktion mit wiederholter Eingabemöglichkeit für denselben Trial zu ermöglichen. In letzterem Falle hätten die Probanden unkontrolliert viele zusätzliche Eingabetrials nutzen können, was die Standardisierung der Steuerungsphasenlänge beeinträchtigt hätte.

seien und ob sie zum Beispiel spezifische Erfahrungen aus der individuellen Arbeitsphase ausgetauscht hätten. Für die frei formulierte Antwort stand ein Eingabefeld zur Verfügung.

Ein weiterer Fragenblock der abschließenden Fragen in der Realbedingung bezog sich auf die wahrgenommene Kompetenz und die subjektive Gewichtung der individuellen Urteile der einzelnen Gruppenmitglieder für das Gruppenurteil. Die Probanden wurden gebeten zu beurteilen, für wie kompetent sie sich und die anderen Gruppenmitglieder im Nachhinein bezogen auf die 60 Steuerungsurteile hielten, die sie in der Gruppenarbeitsphase vorgenommen hatten. Sie sollten sich und die beiden anderen Gruppenmitglieder in eine Rangreihe bringen, indem Sie die drei Aussagen „Das kompetenteste (bzw. das zweitkompetenteste bzw. das am wenigsten kompetente) Mitglied in meiner Gruppe war Mitglied Nummer...“ mit den jeweiligen Gruppenmitgliedsnummern von 1 bis 3 ergänzten. Die relative Kompetenzbeurteilung innerhalb der Gruppe wurde für jedes Gruppenmitglied durch eine absolute Kompetenzeinschätzung auf einer Skala von 1 (*sehr geringe Kompetenz*) bis 7 (*sehr hohe Kompetenz*) ergänzt. Entsprechend wurden die Probanden gefragt, wie sie die Kompetenz der zielgerichteten Systemsteuerung von Mitglied 1 (bzw. Mitglied 2 oder Mitglied 3) auf einer Skala von eins bis sieben beurteilen würden. Anschließend wurden die Probanden gebeten einzuschätzen, wie stark die individuellen Urteile aller drei Gruppenmitglieder in das Gesamturteil der Systemsteuerwerte Ihrer Gruppe eingeflossen seien. Sie sollten dies bitte in Prozent angeben und darauf achten, dass die Summe ihrer Angaben für sie selbst und die beiden anderen Gruppenmitglieder in der Summe 100% ergäben. Zur Veranschaulichung wurde ein Beispiel genannt: Wenn sie den Eindruck gehabt hätten, alle Gruppenmitglieder hätten gleich stark zu dem Gesamturteil beigetragen, dann sollten sie allen drei Gruppenmitgliedern jeweils 33% zuweisen. Wenn sie den Eindruck gehabt hätten, dass ein Gruppenmitglied ganz alleine das Gruppenurteil erstellt hätte, so sollten sie diesem Mitglied 100% und den anderen beiden 0% zuweisen. Die Zuweisung der prozentualen Gewichtung erfolgte durch Vervollständigung der Aussage „Der Anteil von Gruppenmitglied 1 (bzw. Gruppenmitglied 2 oder 3) am Gesamturteil lag bei“. Außerdem sollten die Probanden eingeben, welche Gruppenmitgliedsnummer sie selbst hatten, um eine korrekte Zuordnung und Auswertung der Kompetenzselbsteinschätzung sicherzustellen. Abschließend sollten die Probanden, wie in den beiden Vergleichsbedingungen auch, das Untersuchungsziel benennen.

**Vergütung und Debriefing für alle Bedingungen.** Nach Beantwortung des Abschlussfragebogens wurde allen Probanden auf dem Computerbildschirm mitgeteilt,



dass sie am Ende des Experimentes angekommen seien. Sie wurden gebeten, bitte nicht mit anderen potenziellen Versuchspersonen über die Inhalte dieses Experimentes zu sprechen und ihnen wurde für die Teilnahme gedankt. Die Versuchsleitung, der auf dem Serverrechner angezeigt wurde, an welchem Laptop das Experiment abgeschlossen war und welche Bonuszahlung aus der Systemsteuerung in Phase 2 resultierte, führte die entsprechenden Probanden in einen separaten Experimentalraum. Hier wurde das Versuchspersonenhonorar von fünf Euro Grundvergütung plus der leistungsabhängigen Bonuszahlung, die maximal weitere fünf Euro erreichen konnte, ausgezahlt. Alternativ zu der Grundvergütung wurde eine Versuchspersonenstunde bescheinigt. Der Bonus betrug für mittlere absolute prozentuale Abweichungen kleiner oder gleich 35% (also einem MAPE von .35) 5 Euro, für Abweichungen zwischen 35,1% und 81% 3 Euro, für Abweichungen von 81,1% bis 120% 1 Euro und für Abweichungen größer als 120% 0 Euro.<sup>94</sup> Abschließend wurden die Probanden über das Ziel des Experimentes aufgeklärt, um Verschwiegenheit gebeten und mit Dank für die Teilnahme entlassen.

---

<sup>94</sup> Die Festsetzung der Bonusgrenzen erfolgte gemäß der Leistungsspanne der 30 Probanden im Vortest. Die Bonusgrenze für 5 Euro entsprach der individuellen Bestleistung, die untere Bonusgrenze für 3 Euro entsprach dem 25. Perzentil und die untere Bonusgrenze für 1 Euro entsprach dem 50. Perzentil der MAPE-Werte in der Vorteststichprobe.

## Versuchsleistungsanweisung für Experiment 1

### Ablaufplan für die Versuchsleitung/Instruktionen für die Versuchspersonen

#### 1. Checkliste vor dem Versuch: Vorbereitung

- Benötigtes Material
- Technik überprüfen

#### 2. Einleitende mündliche Instruktion (alle Bedingungen)

„Herzlich willkommen. Schön, dass Sie sich bereit erklärt haben, an der heutigen Untersuchung zum Thema „Systemsteuerungsaufgaben“ teilzunehmen. Uns interessiert, wie Personen individuell oder in der Gruppe diesen bestimmten Aufgabentyp bearbeiten. Die Untersuchung wird circa 45-75 Minuten in Anspruch nehmen. Die Bearbeitungszeit hängt von der Versuchsbedingung und von Ihrem individuellen Arbeitstempo ab.

Das gesamte Experiment findet am Laptop statt. Es gliedert sich in fünf Teile:

- Zunächst werden Sie gebeten einige demographische Angaben zu machen und einige Fragen zu ihrer Person zu beantworten.
- Dann wird Ihnen das zu steuernden Systems schriftlich erläutert und Sie können sich in 20 Übungsdurchgängen mit der Funktionsweise vertraut machen. Es geht um die Steuerung eines Kühlsystems in einem Atomkraftwerk. Dabei handelt es sich um eine Optimierungsaufgabe und das Ziel besteht darin so schnell wie möglich einen Optimalwert zu erreichen und ihn dann möglichst gut beizubehalten. Es sollen also die Temperaturabweichungen vom Optimalwert über alle Steuerungsdurchgänge möglichst gering gehalten werden.
- Als drittes folgt eine individuelle Arbeitsphase, in der Sie über 60 Durchgänge das System nach der bestimmten Zielvorgabe steuern sollen. Nach dieser Phase wird Ihre persönliche Steuerungsleistung erfasst und Sie werden von der Versuchsleitung...
- ...in der vierten Arbeitsphase entweder einer individuellen oder einer Gruppenarbeitsbedingung zugewiesen:
  - [Individualbedingung:] In der individuellen Arbeitsbedingung steuern Sie erneut das System über 60 Durchgänge alleine.
  - [Nominalbedingung:] In einer Gruppenbedingung ist Ihr Computer über einen Server mit zwei anderen Versuchspersonencomputern verbunden. Jedes Gruppenmitglied nimmt räumlich getrennt ohne Absprache mit den anderen 60 individuelle Steuerungsurteile vor, Der Server integriert die Einzeleingaben pro Durchgang zu einem mittleren Gruppenurteil und meldet diesen Gruppenwert und die resultierende Systembeeinflussung nach jedem einzelnen der 60 Durchgänge zurück. Dann erfolgt die Eingabe für den nächsten Durchgang usw.

- [Realbedingung:] In einer zweiten Gruppenbedingung werden Sie in Dreiergruppen auf die Nachbarräume aufgeteilt und sollen für jeden von 60 Steuerungsdurchgängen zunächst ein individuelles Urteil treffen und dann in der Gruppe diskutieren und ein Gruppenurteil festlegen. Für den eingegebenen Gruppenwert erhalten Sie wie gewohnt eine Systemrückmeldung über die resultierende Temperatur.
- Ganz am Ende folgt noch ein allgemeiner Fragebogen, den Sie individuell am PC ausfüllen.
- Die Gruppen werden von der Versuchsleitung (mir) so zusammengestellt, dass sie von der Leistung her möglichst gleich stark sind.
- Für das Experiment gibt es eine Versuchspersonenstunde/5 Euro Grundvergütung und einen leistungsabhängigen Bonus (0 bis 5 Euro), der davon abhängt, wie gut Sie das System nach der Zielvorgabe in der zweiten Steuerungsphase steuern. Bestimmt wird die mittlere prozentuale Abweichung der 60 erreichten Temperaturwerte von der anzustrebenden Optimaltemperatur. Um in der zweiten Phase gut abzuschneiden ist es sehr hilfreich, sich in der ersten individuellen Phase anzustrengen...

Bitte achten Sie darauf, dass Sie mit der Maus immer ganz nach unten scrollen und dann auf den „Weiter-Button“ klicken. Sollte der Button einmal nicht funktionieren, haben Sie wahrscheinlich vergessen, eine Frage zu beantworten/ein Häkchen zu setzen. Wenn Sie offene Fragen nicht beantworten möchten, setzen Sie bitte ein „x“ in das Antwortfeld.

Gibt es jetzt noch Fragen? Dann können Sie beginnen. Viel Spaß!  
Wenn Sie während des Versuches Fragen haben, bitte die Hand heben.“

**3. Datentabelle auf Serverlaptop verfolgen. Wenn abgeschlossen, in Officetabelle kopieren und Leistungskennwerte ablesen** → Gruppeneinteilung möglichst mit 3 unterschiedlichen Fähigkeitsniveaus, z.B. angrenzende Bonusbereiche 0, 1 und 3 (immer mindestens .10 Abweichung zwischen allen paarweisen Vergleichen innerhalb einer Dreiergruppe, also z.B. ein Gruppenmitglied mit einem MAPE von 2.10, eins mit einem MAPE von .99 und eines mit einem MAPE von .50)

MAPE Phase1 $\leq .35$	Bonus = 5
MAPE Phase 1 $> .35$ and $x \leq .81$	Bonus = 3
MAPE Phase 1 $> .81$ and $x \leq 1.24$	Bonus = 1
MAPE Phase 1 $> 1.24$	Bonus = 0

Auf Laptops mit Passwort „test“ Bedingung (I, R, N) auswählen und Gruppennamen vergeben (realx; nominalx bzw. individualxxx ( x ist in den Gruppenbedingungen die Gruppennummer, in der Individualbedingung die Versuchspersonennummer)

Gruppenmitglieder (Realgruppenbedingung) bitten Laptop und Maus mitzunehmen OHNE den Deckel herunterzuklappen (hierdurch geht die Serververbindung verloren), mit Laptops in kleinen Raum führen, Nummern eins bis drei auf Pappschildern zuweisen, wobei die niedrigste Versuchspersonendurchlaufnummer die 1 erhält (also z.B. Proband 170 die 1, 174 die 2, 177 die 3).

#### **4. Mündliche Instruktion:**

[Individualbedingung:] „Sie sollen nun über 60 weitere Zeittakte das Kühlsystem steuern und möglichst den Optimalwert von 4 Grad Celsius erreichen. Sie sind in die individuelle Arbeitsbedingung eingeteilt wurden, in der Sie das System wie bisher allein steuern.

[Nominalbedingung:] Sie sollen nun über 60 weitere Zeittakte das Kühlsystem steuern und möglichst den Optimalwert von 4 Grad Celsius erreichen. Ihr Laptop ist in der zweiten Arbeitsphase über einen Server mit zwei anderen Versuchspersonencomputern verbunden. Jedes Gruppenmitglied dieser virtuellen Dreiergruppe nimmt ohne Absprache mit den anderen 60 individuelle Steuerungsurteile vor, der Server integriert die Einzeleingaben zu einem mittleren Gruppenurteil und meldet diesen Gruppenwert und die resultierende Systembeeinflussung nach jedem einzelnen der 60 Durchgänge zurück. Nach Ihrer ersten Eingabe erscheint so lange „Bitte warten“, bis auch die anderen zwei Gruppenmitglieder ihre Stellradwerte eingegeben haben. Dann wird Ihnen der gemittelte Stellradwert und die daraus resultierende Temperatur zurückgemeldet und Sie können die Eingabe für den folgenden Durchgang durchführen. Nach der 60. Eingabe folgt dann noch ein kurzer individueller Fragebogen.

[Realbedingung:] Sie sollen nun über 60 weitere Zeittakte das Kühlsystem steuern und möglichst den Optimalwert von 4 Grad Celsius erreichen. Für jeden Durchgang überlegen Sie zunächst individuell einen Stellradwert und geben ihn ein. Währenddessen sollten Sie sich nicht beraten oder beim Nachbarn auf den Bildschirm schauen. Wenn Sie alle ihr individuelles Urteil eingegeben haben, tauschen Sie bitte Ihre Vorschläge in folgender Form aus: „Ich habe bei Durchgang 5 den Wert X eingegeben“. Wenn alle drei ihre individuellen Werte genannt haben, können Sie Ihre Ideen über die Funktionsweise des Systems und die optimale Eingabe diskutieren und sollen sich auf einen Gruppenwert einigen, den Sie alle drei in ihre Laptops eingeben. Der individuelle Eingabewert verschwindet, wenn Sie auf „OK“ klicken. Dann können Sie den Gruppenwert eingeben und erhalten wie üblich eine Temperaturrückmeldung auf Basis der eingegebenen Gruppenurteile. Bitte stellen Sie sicher, dass Sie sich immer bei der Eingabe des gleichen Durchganges befinden, indem Sie regelmäßig die Durchgangsnummer nennen.

Falls sich einer in Ihrer Gruppe bei der Eingabe des Gruppenurteils vertippt, sollte er sich für die folgenden Durchgänge an den resultierenden Temperaturen der beiden anderen Gruppenmitglieder orientieren. Einmal getätigte Eingaben können nicht mehr rückgängig gemacht werden.

Dann viel Spaß, ich stehe für Fragen zur Verfügung, falls jetzt oder später noch etwas unklar sein sollte. Nach der 60. Eingabe folgt dann noch ein kurzer individueller Fragebogen.“

**5. Nach Beendigung Daten der zweiten Steuerungsphase in Officetabelle kopieren und Leistungskennwerte/Bonuszahlung ablesen. Honorar auszahlen, Aufklärung über Untersuchungsziel (Lernprozesse in Gruppen, Vergleich der drei Bedingungen...siehe Debriefing-Bogen) und Verabschiedung.**

## **6. Datensicherung**

Daten sofort speichern (Passwort dafür ist ...).

Danach umbenennen der Datei für ein Format, mit dem sie automatisch in SPSS eingelesen werden kann. Dateiname vorne erweitern um Angabe zu Bedingung und – getrennt mit Unterstrichen- die Versuchspersonennummern; Beispiele:

- Real13\_059\_Laborlaptop4\_Dynamische Systeme\_2009070111234.xml
- Nominal01\_060\_Laborlaptop3\_Dynamische Systeme\_20090701112247.xml
- Individual\_099\_Laborlaptop4\_Dynamische Systeme\_20090708113710.xml

### **Debriefing Experiment Dynamische Systeme I**

In diesem Experiment untersuchen wir Gruppenleistung und Lernprozesse bei der Steuerung eines dynamischen Systems.

Bisherige Forschung zeigt, dass Einzelpersonen meistens nicht gut darin sind, Systeme nach bestimmten Zielvorgaben zu steuern bzw. Systemzustände zu kontrollieren. Sie unterliegen vielfältigen Urteilsfehlern und erzielen schlechtere Steuerungsergebnisse als nach einem optimalen Vorgehen oder auch durch Nichteingreifen in das System zu erreichen wären. Auch bei wiederholter Steuerung durchschauen Einzelpersonen Systemeigenschaften wie verzögerte Reaktion des Systems auf Eingriffe oder nichtlineare Beziehungen zwischen Systemvariablen meistens nicht.

Wir gehen der Frage nach, ob Gruppen bei der Steuerung dynamischer Systeme sogenannte „Prozessgewinne“ realisieren können, ob sie also eine bessere Steuerungsleistung erzielen als Individuen. Hierzu vergleichen wir die Leistung real interagierender Gruppen mit zwei weiteren Bedingungen. In der „Individualbedingung“ steuern je drei Probanden das System individuell und die gemittelten Leistungen werden mit der Gruppenleistung verglichen. In der „Nominalbedingung“ sind drei Personen, die individuelle Eingaben machen, über einen Serverrechner vernetzt. Aus ihren drei individuellen Urteilen in jedem Durchgang wird der Mittelwert gebildet, auf den das System dann reagiert. Diese Bedingung dient dazu zu erfassen, welche Vorteile allein durch die statistische Mittelwertbildung von drei Urteilen (ohne Interaktion in der Gruppe) entstehen. Eine darüber hinaus gehende bessere Realgruppenleistung kann dann auf Interaktionsprozesse in der Gruppe zurückgeführt werden.

Darüber hinaus untersuchen wir differenziert, welche Arten von Lernprozessen zur Realgruppenleistung beitragen. Zum einen kann es sein, dass die Gruppenmitglieder durch die Interaktion ihre individuelle Urteilsleistung verbessern (Man nennt dies Group-to-Individual-Transfer). Dies berechnen wir durch einen Vergleich der Individualleistung in den Phasen 1 und 2. Zum anderen kann es sein, dass Koordinationsgewinne auftreten und die Gruppe aus den drei individuellen Urteilsvorschlägen pro Durchgang ein noch besseres Steuerungsurteil abgibt als es der Mittelwert der drei Urteile wäre. Auch dies wird berechnet.

Über die offenen Fragen zu angenommenen Systemzusammenhängen erfassen wir zusätzlich das verbalisierbare Wissen und können auch hier einen Lernzuwachs über die drei Steuerungsphasen berechnen.

### Kodieranweisung zu den Wissensfragen in Experiment 1 und 3

Kodiert werden die Antworten im Datensatz auf Gruppenebene, pro Zeile in SPSS sind nacheinander also die Antworten von drei Versuchspersonen zu finden; die Itemnamen der zweiten Person pro Gruppe haben ein „m2\_“, die der dritten Person ein „m3\_“ vorangestellt. Für jede Versuchsperson wird separat kodiert.

In der SPSS-Datei sind bereits Kodiervariablen zwischen den Antwortspalten eingefügt, bei 0 vs. 1 Kodierungen bitte auch die Null eintragen.

#### Je dreimal pro Zeile zu kodierende Items in chronologischer Reihenfolge:

- **Nach Übungsphase mit 20 Durchgängen ohne Zielvorgabe**
  - **vorgeh\_uephase**  
→ Wie sind Sie in der ersten Explorationsphase des Systems mit den 20 Zeittakten vorgegangen?
  - **annahme\_1**  
→ Wie lauteten Ihre konkreten Annahmen über den Zusammenhang von Stellradeinstellung und Kühltemperatur? Falls sich Ihre Annahmen im Verlauf der Steuerung geändert haben, ordnen Sie sie bitte zeitlich so, dass die aktuell gültige Annahme an letzter Stelle steht. Wenn Sie weniger als 3 Annahmen haben, schreiben Sie bitte "x" in die nicht benötigten Textfelder.
  - **annahme\_2**
  - **annahme\_3**
  
- **Nach der ersten Aufgabenphase mit 60 Durchgängen**
  - **annahme\_1\_phase1**
  - **annahme\_2\_phase1**
  - **annahme\_3\_phase1**
  
- **Nach der zweiten Aufgabenphase mit 60 Durchgängen (individuell oder in der Gruppe)**
  - **annahme\_1\_phase2**
  - **annahme\_2\_phase2**
  - **annahme\_3\_phase2**
  - **vorgehen\_gr (Nur in Realgruppenbedingung ausgefüllt)**  
→ Wie sind Sie bei der Gruppenarbeit vorgegangen? Haben Sie zum Beispiel spezifische Erfahrungen aus der individuellen Arbeitsphase ausgetauscht?
  - **ziel**  
→ Was glauben Sie war das Ziel dieser Untersuchung?
  - **Materialok**  
→ Fanden Sie Teile der Aufgabenstellung und/oder der Fragen unverständlich?
  - **vorgehen (Nur in Individualbedingung ausgefüllt)**

→ Wie sind Sie bei der Bearbeitung vorgegangen? Haben Sie zum Beispiel spezifische Erfahrungen aus der ersten individuellen Arbeitsphase genutzt?

### 1. Kodierung Vorgehen Übungsphase/Phase 2

- (0) = keine Angabe/x/Null
- (1) = Beobachtungsphase (mehrere Durchgänge geschaut, wie das System reagiert; Eigendynamik zu erkennen versucht; über mehrere Durchgänge konstanten Wert eingegeben und abgewartet)
- (2) = Systematische Eingriffsvariation (absichtlich sehr hohen bzw. niedrigen Wert eingegeben und geschaut, wie das System reagiert; mehrmals konstanten)
- (3) = Zeitverzögerte Systemreaktion erkannt und genutzt
- (4) = Vorgehen nach Gefühl/intuitiv
- (5) = falsche Hypothesen/System nicht erkannt, daher kein planmäßiges Vorgehen
- (6) = Ja, Erfahrungen genutzt, aber keine kodierbare Spezifikation

### 2. Kodierung der Systemannahmen (für jedes Annahmefeld in der Übungsphase, Phase 1 und Phase 2)

#### 2.a Richtigkeit der einzelnen Annahmen pro Feld in zwei Spalten kodieren

(1 für vorhanden, 0 für nicht vorhanden kodieren)

- r\_annahme1/2/3\_m1/m2/m3\_ue / ph1/ph2 **Annahme richtig**  
→ (1 für zutreffend, 0 für nicht zutreffend)
- f\_annahme1/2/3\_m1/m2/m3\_ue / ph1/ph2 **Annahme falsch**  
(1 für zutreffend, 0 für nicht zutreffend)  
→ Wenn in einem Annahmefeld eine richtige und eine falsche Annahme stehen mit „1“ in beiden Feldern kodieren

**2.b Inhaltskategorien kodieren** (nach jeder Annahme acht Spalten mit Indizes 0 bis 7; 1 für vorhanden, 0 für nicht vorhanden in jeder Spalte kodieren):

**Itembeispiele, wenn mit „1“ kodiert:**

- annahme1\_1\_m1\_ue : Zusammenhang richtig (1) bei Annahme 1 des ersten Gruppenmitglieds in der Übungsphase
- annahme1\_3\_m1\_ph1 : Schwingendes System vermutet (3) bei Annahme 1 des ersten Gruppenmitglieds in der Phase 1
- annahme2\_7\_m2\_ph2: Zufall angenommen (7) bei Annahme 2 des zweiten Gruppenmitglieds in der Phase 2

→ **(1) Zusammenhang\_richtig: Hohe (niedrige) Stellradeinstellung führt zu hoher (niedriger) Temperatur** (z.B. „Temperatur sinkt, wenn Stellrad nach unten gedreht wird.“; „Große Veränderungen führen zu stärkeren Temperaturänderungen.“)

→ **(2) Zeitverzögerte Reaktion des Systems richtig** (z.B. „Temperatur reagiert erst später“; „System ist träge“)



- **(3) „sinusförmige“ Schwingung richtig** (z.B. „Temperaturschwankungen sind auch systembedingt“, „nicht-linear“)
- **(4) Zusammenhang\_falsch** (z.B. hohe Stellradeinstellung senkt Temperatur, Zahlen im mittleren Bereich erhöhen die Temperatur)
- **(5) Zeitverzögerte Reaktion des Systems falsch** (z.B. „System reagiert sofort auf Eingaben“)
- **(6) Magisches Zahlendenken** (z.B. „1 = hoch, 0 = runter“)
- **(7) Zufall/kein schlüssiges System** (z.B. „willkürlich?“)
- **(0) optimale Stellradeinstellung zur Systemstabilisierung richtig:** (23 ist korrekt, Angaben zwischen Einstellung 18 bis 30 als richtig kodieren, z.B. „Die Stellradeinstellungen 25-29 erzeugen in etwa eine Temperatur von 4°C.“)

→ Falls eine Aussage unklar/nicht schlüssig für den Kodierer ist, unter diffusem Systemverständnis mit (7) kodieren!

→ Wenn pro Annahmefeld mehrere Annahmen aufgeschrieben werden, sollten beide kodiert werden, also z.B. für annahme1\_ue (1) und (2) mit Einserkodierung in den Spalten annahme1\_1\_m1\_ue und annahme1\_2\_m1\_ue, entsprechend bei Kodierung richtig/falsch auch als zwei Annahmen berücksichtigen

→ Wenn Annahmen innerhalb einer Arbeitsphase (z.B. Annahmen 1 bis 3 der Phase 1) mehrfach genannt werden, nur beim ersten Mal kodieren).

→ Kontrolle vollständiger Kodierung: Anzahl der gesamten Annahmen gleich Anzahl der Inhaltskodierungen pro Block von drei Annahmen!

### 3. Kodierung Ziel

- (0) = falsche/keine Zielangabe
- (1) = richtig: Vergleich individueller und Gruppenleistung
- (2) = richtig: Leistung/Lernfortschritt bei der Systemsteuerung

### 4. Kodierung Materiallok

„Waren Teile der Instruktionen/Fragen unverständlich?“

- (0) = nein (nö, x, -, ...)
- (1) = ja, doppelte Verneinung in Fragen (NEO-FFI gemeint)
- (2) = ja, sonstige Kritik an Verständlichkeit der Frageitems oder Instruktionen (Materialkritik am „neuen“ Fragebogenmaterial)
- (3) = Aufgabe zu schwer, Zusammenhänge des Systems unverständlich etc. (bezieht sich auf Bewältigungsfähigkeit, nicht auf Verständlichkeit der Aufgabenstellung)

→ Auffällige Angaben mit Probanden-ID gesondert notieren

## Ergänzende Analysen zu Lerneffekten auf individueller Ebene in Experiment 1

**Analysen zu Lerneffekten auf individueller Ebene.** In Abschnitt 4.4.3 wurde das Auftreten von I-I Transfer in der Individualbedingung und von individueller Leistungsverbesserung in der Realbedingung und in der Nominalbedingung untersucht. Hierzu wurden die Ergebnisse einer 3 (Bedingung: Realbedingung vs. Nominalbedingung vs. Individualbedingung) x 3 (Gruppenmitglied: bestes vs. mittleres vs. schlechtestes) Kovarianzanalyse (ANCOVA) mit der Bedingung als Zwischensubjektfaktor, dem Gruppenmitglied als Innersubjektfaktor und der Startkompetenz der Gruppen als Kovariate berechnet. In diesem Abschnitt erfolgt eine genauere Analyse des individuellen Fertigkeitsszuwachses in Abhängigkeit vom Status der Gruppenmitglieder, indem jede Bedingung separat betrachtet wird.

Für die Übungseffekte in der Individualbedingung zeigte eine ANCOVA mit dem Gruppenmitglied als Innersubjektfaktor (bestes, mittleres, schlechtestes) und der Startkompetenz der Gruppen als Kovariate keinen signifikanten Haupteffekt für das Gruppenmitglied,  $F(2, 66) = .08, p = .925, \eta^2_p = .00$ . Der Effekt der Kovariate Startkompetenz war hingegen signifikant,  $F(1, 33) = 8.77, p = .006, \eta^2_p = .21$ , da Mitglieder in Gruppen mit Startkompetenz größeren individuellen Lerntransfer zeigten,  $r(105) = .46, p = .006$ , zweiseitige Testung. Es lag kein signifikanter Interaktionseffekt zwischen der Startkompetenz und dem Faktor Gruppenmitglied vor,  $F(2, 66) = .42, p = .662, \eta^2_p = .01$ . Paarweise Bonferroni-korrigierte post hoc-Tests zeigten nur beim des schlechtesten ( $M = .34, SD = .56$ ) mit dem besten Gruppenmitglied ( $M = .05, SD = .46$ ) einen signifikanten Unterschied im Ausmaß des I-I Transfers,  $t(34) = .29, p = .022, d = 0.52$ . Der Unterschied in der individuellen Leistungsverbesserung zwischen dem mittleren ( $M = .24, SD = .53$ ) und dem besten, sowie zwischen dem schlechtesten und dem mittleren Gruppenmitglied waren nicht signifikant,  $t(34) = .19, p = .344, d = 0.38$  bzw.  $t(34) = .10, p = 1, d = 0.17$ . Für die Realbedingung sowie für die Nominalbedingung wurden ebenfalls separate ANCOVAS mit dem Innersubjektfaktor Gruppenmitglied und der Startkompetenz als Kovariate durchgeführt. Entsprechend der in Tabelle 14 ersichtlichen (und im Vergleich zur Individualbedingung relativ großen) Mittelwertunterschiede waren alle Bonferroni-korrigierten Paarvergleiche zwischen den Gruppenmitgliedern in diesen Bedingungen mit  $p < .001$  signifikant.

Durchgeführte post-hoc Einzel-*t*-Tests gegen Null zeigten, dass das mittlere und schlechteste Gruppenmitglied in der Individualbedingung ihre individuelle Leistung in der Steuerungsphase 2 signifikant verbesserten. Das Gruppenmitglied mit der besten Individualleistung in Phase 1 zeigte hingegen keinen signifikanten Leistungszuwachs in der Steuerungsphase 2. Die genauen *t*-Werte, *p*-Werte und die Effektgrößen *d* für das jeweils beste, mittlere und schlechteste Gruppenmitglied in der Individual- sowie in der Nominal- und die Realbedingung sind in Tabelle 14 in der Übersicht dargestellt.

Tabelle 14: *Differenzmaßwerte und Ergebnisse separater Einstichproben-t-Tests gegen Null für den individuellen Leistungszuwachs in der Steuerungsphase 2 des Experiments 1*

Bedingung	<i>M</i> ( <i>SD</i> )	<i>t</i> (34)	<i>p</i>	Cohen's <i>d</i>
Real				
bestes Mitglied	-.11 (.32)	-2.02	.051	0.34
mittleres Mitglied	.49 (.42)	6.88	< .001	1.17
schlechtestes Mitglied	1.07 (.52)	12.26	< .001	3.30
gesamt	.48 (.37)	7.81	< .001	1.32
Nominal				
bestes Mitglied	-.12 (.45)	1.64	.111	0.28
mittleres Mitglied	.37 (.40)	5.44	< .001	0.92
schlechtestes Mitglied	.99 (.63)	9.24	< .001	1.56
gesamt	.41(.44)	5.55	< .001	0.94
Individual				
bestes Mitglied	.05 (.46)	.61	.548	0.10
mittleres Mitglied	.24 (.53)	2.66	.012	0.45
schlechtestes Mitglied	.34 (.56)	3.55	.001	0.60
gesamt	.21 (.37)	3.31	.002	0.56

Auch in der Real- und in der Nominalbedingung verbesserte sich das beste Mitglied nicht signifikant und erbrachte deskriptiv sogar eine schlechtere Leistung als in der Steuerungsphase 1, *p* = .051 bzw. *p* = .111. Das mittlere sowie das schlechteste Mitglied

zeigten in der Realbedingung und in der Nominalbedingung einen noch deutlicheren Leistungszuwachs als in der Individualbedingung, alle  $p < .001$ . Der über alle Gruppenmitglieder hinweg insgesamt deutliche Leistungszuwachs in der Real- und in der Nominalbedingung mit jeweils  $p < .001$ , der den Leistungszuwachs durch reine Übungseffekte in der Individualbedingung mit  $p = .002$  übertrifft, basiert also auf einer stark verbesserten Individualleistung des mittleren und des schlechtesten Gruppenmitglieds.

## **Explorative Analyse: Zusammenhang von Persönlichkeitsfaktoren mit der Steuerungsleistung und der Gewichtung der individuellen Urteile**

Alle Probanden in Experiment 1 beantworteten vor Beginn der Systemsteuerung die 60 Items des Persönlichkeitsinventars NEO-FFI von Borkenau und Ostendorf (1993). Den Probanden wurde am Monitor die vollständige Instruktion sowie die Erläuterung der fünfstufigen Skala zur Selbstbeschreibung der eigenen Person angezeigt. Per Mausclick konnten die Probanden für jedes Item auf einer fünfstufigen Skala von -2 (*Starke Ablehnung*) bis 2 (*Starke Zustimmung*) den Zustimmungsgrad auswählen, der am ehesten auf sie zutraf. Gemäß der Auswertungsinstruktionen wurde ein Teil der Itemantworten invertiert und es wurden anschließend für jeden Probanden die Mittelwerte aus je 12 Itemantworten für die Persönlichkeitsdimensionen Neurotizismus, Extraversion, Verträglichkeit und Gewissenhaftigkeit berechnet.

Anschließend wurden mögliche Zusammenhänge zwischen jeder der fünf Persönlichkeitsdimensionen und der individuellen Leistung in der ersten Steuerungsphase überprüft. Für den Persönlichkeitsfaktor Neurotizismus zeigte sich ein signifikanter positiver Zusammenhang mit dem MAPE für Phase 1,  $r(315) = .15$ ,  $p = .007$ , zweiseitige Testung. Eine höhere Ausprägung in Neurotizismus ging also tendenziell mit einer schlechteren Steuerungsleistung einher. Die verbleibenden vier Persönlichkeitsfaktoren wiesen keine signifikanten Zusammenhänge mit der individuellen Steuerungsleistung auf, alle  $r(315) < |.09|$ , alle  $p > .05$ .

Als nächstes wurde für die Realbedingung der Zusammenhang zwischen der Ausprägung der NEO-FFI-Faktoren auf Gruppenebene und der Steuerungsleistung in Phase 2 analysiert. Hierzu wurde für jedes der fünf Persönlichkeitsmerkmale jeweils der Mittelwert der Ausprägung von je drei Gruppenmitgliedern berechnet und mit dem MAPE als Gruppenleistungsmaß korreliert. Keine der Korrelationen zwischen den durchschnittlichen Merkmalsniveaus der Gruppe das und der Gruppenleistung war signifikant,  $r(35) < |.30|$ , alle  $p > .05$ .

Ausgehend von bisherigen empirischen Befunden zur Gruppenleistung ist insbesondere das Merkmal Extraversion für diese Arbeit relevant. Der Trait Extraversion misst Eigenschaften mit interpersonalem Bezug wie Aktivität und Dominanz, die sich auf die Qualität sozialer Interaktion auswirken (McCrae, & Costa, 1992). Littlepage, Schmidt, Whisler und Frost (1995) fanden einen signifikanten Zusammenhang zwischen der

Extraversion und dem Redeanteil von Personen in Problemlösegruppen. Forschung zu „emergent leadership behaviors“ hat gezeigt, dass Extraversion ein guter Prädiktor dafür ist, ob eine Person in einer führerlosen Gruppe, die eine Problemlöseaufgabe bearbeitet, die Führungsrolle übernimmt und maßgeblich die Interaktion gestaltet (Kickul & Neuman, 2000). Bisherige Forschung hat darüber hinaus aufgezeigt, dass Gruppen bei mehrdeutigen visuellen Entscheidungsaufgaben (Bonner, 2000) und bei quantitativen Schätzaufgaben mit eingeschränkter Demonstrierbarkeit der richtigen Lösung die Beiträge extravertierterer Gruppenmitglieder stärker bei der Lösungsfindung gewichten als die Beiträge der weniger extravertierten Mitglieder (Bonner & Sillito, 2011; Bonner, Sillito, & Baumann, 2007). Eine niedrige Demonstrierbarkeit und ein starker Einfluss des extravertiertesten Gruppenmitglieds war in diesen Studien dann gegeben, wenn es keine Hinweise auf die richtige Lösung in Form einer Mehrheitsmeinung (Bonner, 2000), in Form von Expertisefeedback oder einer *bridge-building*-Intervention (Bonner & Sillito, 2011) oder in Form von Referenzwertinformationen gab (Bonner, Sillito, & Baumann, 2007). Bei Steigerung der Demonstrierbarkeit der richtigen Lösung durch die genannten Maßnahmen entsprach die Gruppenlösung hingegen eher der Majoritätsmeinung bzw. einer Gewichtung der Mitgliederbeiträge nach Akkuratheit und Expertise.

Bei der dynamischen Systemsteuerungsaufgabe in Experiment 1 handelt es sich wie unter 4.3 beschrieben, um eine intransparente, komplexe Aufgabe ohne unmittelbare Demonstrierbarkeit der richtigen Lösung bzw. ohne perfekte Demonstrierbarkeit der Lösungsstrategie über mehrere Trials. Daher ist es erstens plausibel anzunehmen, dass die Gruppenmitglieder von der Ausprägung extravertierter Verhaltensweisen während der Gruppeninteraktion auf die Expertise der anderen Mitglieder geschlussfolgert haben und zweitens deren Individualurteile stärker bei der Gruppenurteilsfindung gewichtet haben.

In weiteren Analysen wurde untersucht, ob es einen Zusammenhang zwischen der Selbsteinschätzung in Extraversion und der von den Gruppenmitgliedern wahrgenommenen Expertise sowie wahrgenommenen und tatsächlichen Gewichtung der Beiträge eines Gruppenmitglieds gab. Zunächst wurden die drei Mitglieder jeder Gruppe entsprechend ihrer Extraversionsratings im NEO-FFI in eine Rangreihe gebracht. Ob die Ausprägung an Extraversion zu unterschiedlich stark wahrgenommener Expertise der Gruppenmitglieder führt, wurde in einer einfaktoriellen MANOVA mit dem Innersubjektfaktor Extraversionsrang (höchste vs. mittlere vs. niedrigste Extraversionsausprägung pro Gruppe) und der über die drei Gruppenmitglieder gemittelten subjektiven

Kompetenzeinschätzung als abhängige Variable überprüft.<sup>95</sup> Die durchschnittliche Kompetenzzuschreibung unterschied sich nicht signifikant nach Extraversionsausprägung der Gruppenmitglieder,  $F(2,64) = 0.33$ ,  $p = .722$ ,  $\eta^2_p = .01$ . Die Kompetenzbeurteilung des jeweils extravertiertesten Gruppenmitglieds ( $M = 4.89$ ,  $SD = .95$ ) war nahezu identisch mit der Kompetenzbeurteilung des mittelmäßig extravertierten ( $M = 4.99$ ,  $SD = .95$ ) sowie des am wenigsten extravertierten Gruppenmitglieds ( $M = 5.08$ ,  $SD = 1.05$ ).

Obwohl die Realgruppen extravertiertere Mitglieder *nach* der gemeinsamen Steuerungsphase nicht signifikant kompetenter einschätzten, ist es denkbar, dass sie deren Individualurteile aufgrund des vermeintlich höheren Redeanteils oder selbstbewussteren Auftretens während der Interaktion stärker gewichteten. In einer zweiten MANOVA mit dem Innersubjektfaktor Extraversionsrang (höchste vs. mittlere vs. niedrigste Extraversionsausprägung pro Gruppe) und der über die drei Gruppenmitglieder gemittelten subjektiven Gewichtung des Gruppenmitglieds als abhängige Variable<sup>96</sup> wurde exploriert, ob die Probanden in der Realbedingung einen stärkeren Einfluss der extravertierteren Gruppenmitglieder auf die Gruppenurteile wahrnahmen. Die subjektive Gewichtung der individuellen Beiträge unterschied sich nicht signifikant nach Extraversionsausprägung der Gruppenmitglieder,  $F(2,64) = .61$ ,  $p = .545$ ,  $\eta^2_p = .02$ . Sie betrug beim extravertiertesten Gruppenmitglied ( $M = 34.53$ ,  $SD = 8.54$ ), beim mittelmäßig extravertierten ( $M = 31.73$ ,  $SD = 7.93$ ) und bei dem Gruppenmitglied mit relativ gering ausgeprägter Extraversion ( $M = 33.48$ ,  $SD = 8.96$ ).

Im Folgenden wurde die tatsächliche Gewichtung nach Extraversion anhand der Passung eines *Extroversion Models auf die Gruppenleistungsdaten* überprüft. Dieses Modell geht von einer vollständigen Gewichtung der Individualurteile des extravertiertesten Gruppenmitglieds aus.<sup>97</sup> Eine 2 (Leistungsmaß: Realgruppenleistung vs. Extroversion

---

<sup>95</sup> Jedes Realgruppenmitglied wurde nach der Gruppensteuerungsphase gebeten, die Kompetenz der zielgerichteten Systemsteuerung von Mitglied 1 (bzw. Mitglied 2 oder Mitglied 3) auf einer Skala von 1 bis 7 zu beurteilen. Die abhängige Variable der subjektiven Kompetenzeinschätzung wurde durch Mittelung der Einschätzungen der drei Gruppenmitglieder (einer Selbsteinschätzung und zwei Fremdeinschätzungen) bestimmt. Das Ergebnis der ANOVA änderte sich nicht, wenn die abhängige Variable der subjektiven Kompetenz nur auf Basis der Fremdurteile der jeweils verbleibenden zwei Gruppenmitglieder berechnet wurde.

<sup>96</sup> Die Probanden schätzten nach der Realgruppeninteraktion ein, wie stark die individuellen Urteile pro Gruppenmitglied über alle Trials hinweg in das Gesamturteil der Systemsteuerwerte Ihrer Gruppe eingeflossen seien.

<sup>97</sup> Diese extremste Form der extraversionsabhängigen Gewichtung wurde gewählt, um das Extroversion Model besser vom Average Model differenzieren zu können. Darüber hinaus wäre eine

Model) x 4 (Trialblock: 1 bis 4) ANOVA mit beiden Faktoren als Innersubjektfaktoren zeigte keinen signifikanten Haupteffekt für die Variable Leistungsmaß,  $F(1, 34) = 0.09$ ,  $p = .761$ ,  $\eta^2_p = .00$ . Es zeigte sich ein signifikanter Haupteffekt des Trialblocks,  $F(1.53, 51.87) = 49.15$ ,  $p < .001$ ,  $\eta^2_p = .59$ , der durch einen abfallenden Trend der MAPE-Werte von Block 1 bis Block 3 charakterisiert war und sowohl durch einen signifikanten linearen als auch durch einen quadratischen Innersubjektcontrast beschrieben werden konnte,  $F(1, 34) = 53.34$ ,  $p < .001$ ,  $\eta^2_p = .61$  bzw.  $F(1, 34) = 62.29$ ,  $p < .001$ ,  $\eta^2_p = .65$ . Die MAPE-Werte pro Trialblock waren für beide Modelle bis zur zweiten Nachkommastelle identisch und sanken vom ersten ( $M = 1.38$ ,  $SD = .47$ ) über den zweiten ( $M = .68$ ,  $SD = .52$ ) bis zum dritten Trialblock ( $M = .47$ ,  $SD = .51$ ) ab und blieben im vierten Trialblock auf einem guten Leistungsniveau ( $M = .48$ ,  $SD = .58$ ).<sup>98</sup> Zwischen Trialblock und Leistungsmaß lag keine signifikante Interaktion vor,  $F(1.51, 51.30) = 0.17$ ,  $p = .784$ ,  $\eta^2_p = .01$ . Der RMSD des Extraversion Models betrug  $.65$  ( $SD = .54$ ). Die Passung des Extraversion Models wurde mit dem RMSD des Average Models von  $.03$  verglichen, das von den weiteren untersuchten normativen Modellen den besten Fit bezüglich der Realgruppenleistungsdaten aufwies. Ein Wilcoxon-Test zeigte, dass der RMSD des Average Models signifikant niedriger ausfiel als der des Extraversion Models,  $Z = 5.16$ ,  $p < .001$ . Für jede der 35 Realgruppen passte das Average Model besser auf die Gruppenleistungsdaten. Deskriptiv entsprach die Passung des Extraversion Models in etwa der des Initial Best Member Models mit  $.64$  und übertraf die Passung des Accuracy Models ( $.80$ ) sowie des True Best Member Models ( $.82$ ).

Zusammenfassend kann festgehalten werden, dass die relative Extraversionausprägung eines Gruppenmitglieds in der Realbedingung keine Unterschiede in der Kompetenzeinschätzung nach sich zog. Konsistent zu dieser extraversionunabhängigen Kompetenzeinschätzung unterschied sich die berichtete Gewichtung der Individualurteile des extravertiertesten, mittelmäßig extravertierten und am wenigsten extravertierten Gruppenmitgliedes für die Festlegung der Gruppenurteile nicht

---

relatives Gewichtungsmodell in Abhängigkeit der relativen Extraversionausprägung innerhalb einer Realgruppe [wie es Bonner et al. (2007) für Schätzaufgaben verwendeten] für die hier durchgeführten Berechnungen zur dynamischen Systemsteuerungsaufgabe eher ungeeignet, da bei letzteren der Leistungsoutput nicht direkt proportional zum gewichteten Urteilsinput ist.

<sup>98</sup> Ein Wilcoxon-Vorzeichen-Rang-Test mit der Summe der absoluten prozentualen Abweichungen über alle Trials der Steuerungsphase 2 als abhängiger Variable zeigte, dass die Realgruppenleistung dem Leistungsniveau des extravertiertesten Gruppenmitglieds entsprach,  $z = -.48$ ,  $p = .635$ , zweiseitige Testung: Demnach erzielten 16 der 35 Realgruppen größere Abweichungen in Bezug auf die Zieltemperatur als das extravertierteste Mitglied ihrer Gruppe.



voneinander. Das Average Model, das eine gleichmäßige Gewichtung aller Gruppenmitglieder annimmt, beschrieb die Realgruppenleistung signifikant besser als ein Extroversion Model mit einer vollständigen Gewichtung des extravertiertesten Mitglieds.

Doch hätte sich eine extraversionsbasierte Gewichtungsstrategie der Realgruppen positiv auf die Leistung ausgewirkt? Bei vollständiger Gewichtung des extravertiertesten Gruppenmitglieds hätte ein MAPE von .75 ( $SD = .44$ ) resultiert, der nicht signifikant vom MAPE des Gruppenpotenzials mit .74 ( $SD = .42$ ) verschieden gewesen wäre,  $t(34) = 0.41$ ,  $p = .685$ ,  $d = 0.02$  zweiseitige Testung. Die Ausrichtung der Systemsteuerung am extravertiertesten Mitglied hätte also über alle Gruppen hinweg keine Koordinationsverluste nach sich gezogen, aber auch keine generellen Koordinationsgewinne ermöglicht. Bei 40% der Realgruppen war das extravertierteste Gruppenmitglied zugleich das individuell beste Gruppenmitglied während der zweiten Steuerungsphase. Für diese Teilmenge der Stichprobe hätte eine Festlegung der Gruppenurteile nach Vorschlag des extravertiertesten Mitglieds Koordinationsgewinne ermöglicht.<sup>99</sup> In den verbleibenden 60% der Fälle wäre es hingegen leistungsförderlicher gewesen, nach Expertise und nicht nach Extraversion zu gewichten.

---

<sup>99</sup> Ein  $t$ -Test für abhängige Stichproben zeigte im Vergleich zum Gruppenpotenzial eine signifikant bessere Steuerungsleistung auf Basis des Extroversion Models,  $t(34) = 5.41$ ,  $p < .001$ .

## Explorative Analyse: Zusammenhang von motivationalen Variablen und Steuerungsleistung in Experiment 1

Nach der Explorationsphase, der individuellen ersten Steuerungsphase und der zweiten Steuerungsphase im Individual-, Nominal- oder Realgruppenkontext machten die Probanden jeweils individuell Angaben zu ihrem Wohlbefinden und ihrer Aufgabenmotivation. Erstens wurden die Probanden gefragt, wie wohl sie sich während der Abgabe der letzten 20 (in der Explorationsphase) bzw. 60 Systemsteuerungsurteile gefühlt hätten. Hierbei war eine Einstufung von 1 = *sehr unwohl* bis 7 = *sehr wohl* möglich. Zweitens sollten die Probanden einschätzen, wie wichtig es Ihnen sei, das System der Temperatursteuerung zu verstehen (Einstufung von 1 = *sehr unwichtig* bis 7 = *sehr wichtig*). Drittens sollten sie mit einer Einstufung von 1 = *sehr wenig Mühe* bis 7 = *sehr viel Mühe* angeben wie viel Mühe sie sich während der Abgabe der 20 bzw. 60 Steuerungsurteile während der Systemexploration bzw. Systemsteuerung gegeben hatten. Da die Angaben der drei Gruppenmitglieder in der Nominal- und Realbedingung nach der computervermittelten bzw. face-to-face Interaktion in der zweiten Steuerungsphase nicht unabhängig voneinander waren, wurde für jede der drei abhängigen Variablen ein Gruppenmittelwert berechnet. Jede der drei Variablen wurde in einer MANOVA mit einem 3 (Innersubjektfaktor Steuerungsphase: Explorationsphase vs. Phase 1 vs. Phase 2) x 3 (Bedingung: Real vs. Nominal vs. Individual) analysiert.

Für die abhängige Variable *Wohlbefinden* zeigte sich ein signifikanter Unterschied im Phasenverlauf,  $F(2, 204) = 80.61$ ,  $p < .001$ ,  $\eta^2_p = .44$ . Dieser wurde durch einen linearen sowie einen quadratischen Innersubjektcontrast qualifiziert, beide  $p < .001$ , beide  $F(1, 102) > 40.03$  und  $\eta^2_p = .28$ . Wie in Abbildung 19 zu sehen ist, stieg das Wohlbefinden von der zweiten zur dritten Phase in allen Bedingungen an. Auch der Haupteffekt für die Bedingung war mit  $F(2, 102) = 3.60$ ,  $p = .031$  und  $\eta^2_p = .07$  signifikant. Bonferroni-adjustierte Einzelvergleiche dass dieser Effekt auf einem signifikant größeren Wohlbefinden in der Realbedingung als in der Individualbedingung basierte,  $p = .026$ . Die anderen paarweisen Vergleiche waren beide mit  $p = .572$  nicht signifikant. Desweiteren lag eine signifikante Interaktion von Phase x Bedingung vor,  $F(2, 204) = 8.62$ ,  $p < .001$ ,  $\eta^2_p = .145$ . Das subjektive Wohlbefinden der Realgruppenmitglieder nahm von der ersten zur zweiten Steuerungsphase deutlich stärker zu als das Wohlbefinden der Probanden in der Nominal- und in der Individualbedingung. Dies könnte zum einen mit der entlastenden und unterstützenden Situation einer interaktiven face-to-face Aufgabenbearbeitung

zusammenhängen. Zum anderen deutet eine signifikante Korrelation zwischen der gruppenbasierten Steuerungsleistung in Phase 2 und dem mittleren Wohlbefinden auf Gruppenebene darauf hin, dass die Probanden nach einer relativ erfolgreichen Systemsteuerung ein größeres Wohlbefinden berichteten,  $r(105) = -.37$ ,  $p < .001$ , zweiseitige Testung.

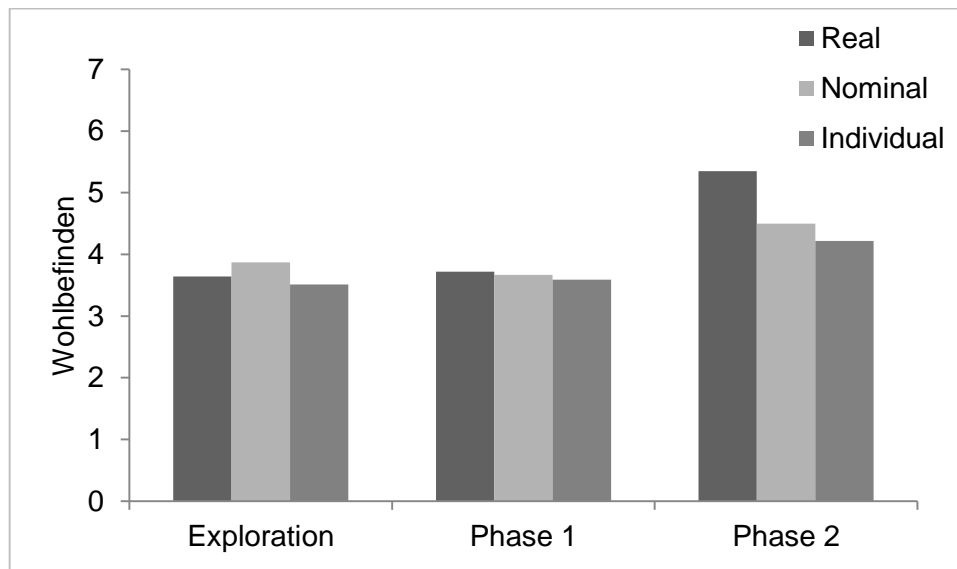


Abbildung 19. Subjektives Wohlbefinden in der Explorationsphase, der Steuerungsphase 1 und der Steuerungsphase 2 getrennt nach Bedingungen in Experiment 1.

Die zweite Variable stellte einen Indikator der Aufgabenmotivation dar und erfasste die subjektiv wahrgenommene *Wichtigkeit, das dynamische System zu verstehen*. Die individuellen Ratings der Probanden wurden wieder zu Gruppenmittelwerten verrechnet und gingen als abhängige Variable in eine 3 (Innersubjektfaktor Steuerungsphase: Explorationsphase vs. Phase 1 vs. Phase 2) x 3 (Bedingung: Real vs. Nominal vs. Individual) MANOVA ein. Der Haupteffekt für die Steuerungsphase war mit  $F(2, 204) = 80.61$ ,  $p < .001$ ,  $\eta^2_p = .44$  signifikant und wurde durch einen quadratischen Innersubjektcontrast beschrieben,  $p < .001$ ,  $F(1,204) = 12.96$ ,  $\eta^2_p = .11$ . Die Motivation, das Kühlsystem zu verstehen unterschied sich insgesamt nicht signifikant zwischen den Bedingungen,  $F(2, 102) = .06$ ,  $p = .939$ ,  $\eta^2_p = .00$ . Allerdings war die Interaktion zwischen der Steuerungsphase und Bedingung signifikant,  $F(2, 204) = 8.616$ ,  $p < .001$ ,  $\eta^2_p = .15$ . Während die Wichtigkeit, ein gutes Systemverständnis zu erlangen in der Individual- und der Nominalbedingung von der ersten zur zweiten Steuerungsphase abnahm, hielt sich die Verständnismotivation wie in Abbildung 20 ersichtlich in der Realbedingung in der

Gruppenphase auf dem Niveau der vorangehenden individuellen Steuerungsphase. Während der Steuerungsphase 2 unterschied sich die subjektive Wichtigkeit des Systemverständnisses nicht signifikant zwischen den Bedingungen,  $F(2, 102) = 0.84$ ,  $p = .435$ ,  $\eta^2_p = .02$ . Deskriptiv lag es (anders als während der Explorationsphase und der individuellen Steuerungsphase) in der Realbedingung im Mittel geringfügig höher ( $M = 5.74$ ,  $SD = .98$ ) als in der Nominalbedingung ( $M = 5.58$ ,  $SD = .83$ ) und der Individualbedingung ( $M = 5.46$ ,  $SD = .97$ ).

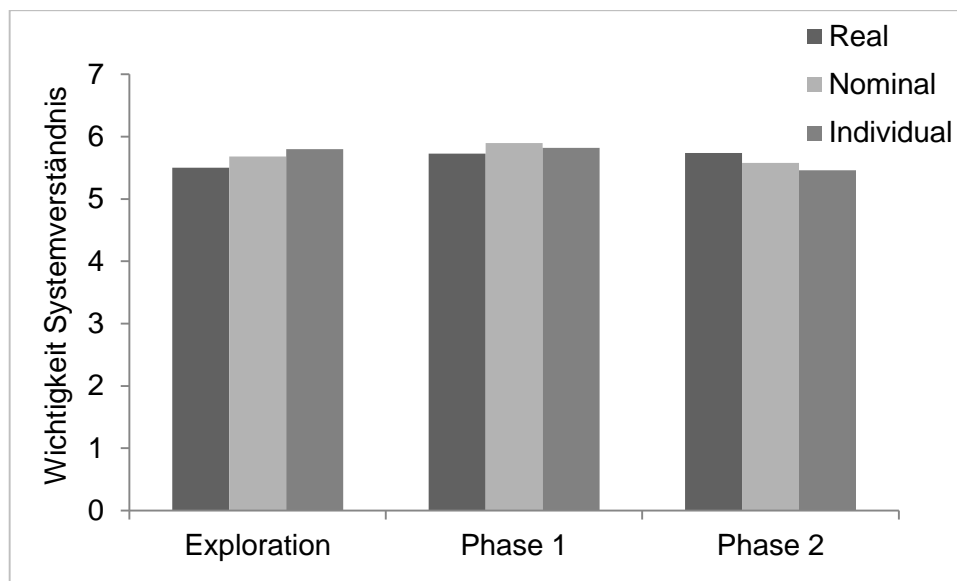


Abbildung 20. Subjektive Wichtigkeit, das System zu verstehen in der Explorationsphase, der Steuerungsphase 1 und der Steuerungsphase 2 getrennt nach Bedingungen in Experiment 1.

Die dritte erhobene Variable stellte einen zweiten Indikator für die Aufgabenmotivation dar und fragte nach der *subjektiven Anstrengung* in Form der Einschätzung der Mühe, die sich die Probanden bei der Systemsteuerung in der jeweils vorangehenden Phase gegeben hatten. Für die auf Gruppenebene gemittelten Ratings wurde abermals eine 3 (Innersubjektfaktor Steuerungsphase: Explorationsphase vs. Phase 1 vs. Phase 2) x 3 (Bedingung: Real vs. Nominal vs. Individual) MANOVA berechnet. Der Haupteffekt für die Steuerungsphase war auch hier signifikant,  $F(2, 204) = 19.33$ ,  $p < .001$ ,  $\eta^2_p = .16$ , und konnte durch einen linearen Innersubjektcontrast erklärt werden,  $F(1, 204) = 33.43$ ,  $p < .001$ ,  $\eta^2_p = .25$ . Die Bedingungen unterschieden sich nicht signifikant im Ausmaß der gegebenen Mühe,  $F(2, 102) = 0.31$ ,  $p = .734$ ,  $\eta^2_p = .01$ . Auch die Interaktion zwischen

Steuerungsphase und Bedingung war knapp nicht signifikant,  $F(2, 204) = 2.39$ ,  $p = .05$ ,  $\eta^2_p = .052$ . Abbildung 21 zeigt, dass die gegebene Mühe in der Real- und der Nominalbedingung linear über die drei Steuerungsphasen anstieg. In der Individualbedingung nahm der Motivationsindikator von der Explorationsphase zur ersten individuellen Steuerungsphase ebenfalls zu, aber in der zweiten individuellen Steuerungsphase im Mittel wieder leicht ab. Die in einer separaten Analyse betrachtete subjektive Mühe während der *zweiten* Steuerungsphase unterschied sich nicht signifikant zwischen den Bedingungen,  $F(2, 102) = 2.66$ ,  $p = .075$ ,  $\eta^2_p = .05$ .

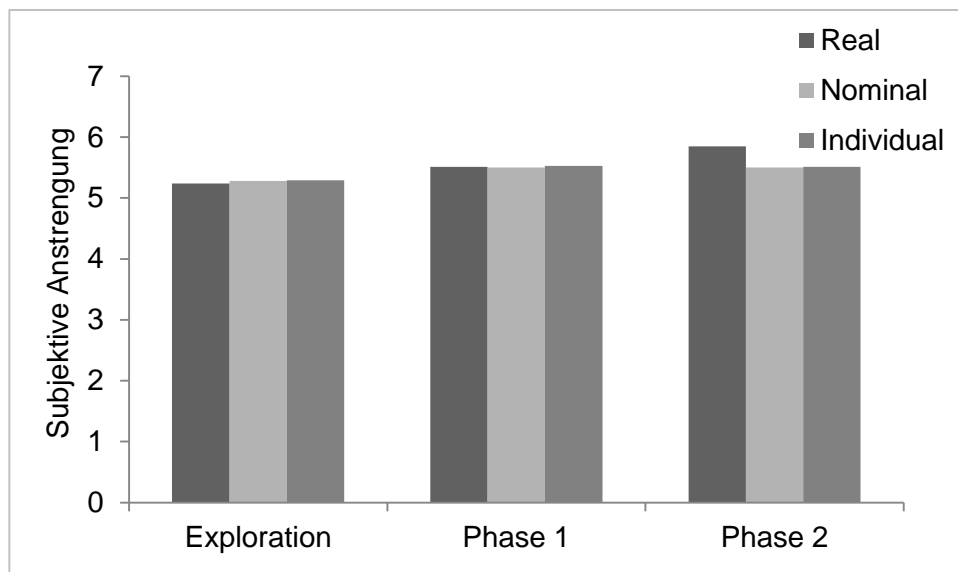


Abbildung 21. Subjektive Anstrengung bei der Systemexploration bzw. -steuerung in der Explorationsphase, der Steuerungsphase 1 und der Steuerungsphase 2 getrennt nach Bedingungen in Experiment 1.

Die beiden subjektiven Motivationsmaße waren deskriptiv in der Realbedingung während der zweiten Steuerungsphase am stärksten ausgeprägt (vgl. Abbildungen 20 und 21). Lässt dies die Schlussfolgerung zu, dass die deskriptiv größere Leistungszunahme von der ersten zur zweiten Steuerungsphase in der Realbedingung gegenüber der Nominal- und Individualbedingung auf eine gesteigerte Motivation im Gruppenkontext zurückzuführen ist? Eine nicht signifikante Korrelationen zwischen dem gruppenbasierten MAPE in Steuerungsphase 2 und dem ersten Motivationsmaß spricht gegen eine Leistungssteigerung aufgrund einer erhöhten Motivation das System zu verstehen [ $r(105) = -.11$ ,  $p = .274$ ]. Allerdings korrelierte die subjektiv gegebene Mühe mit der Steuerungsleistung in Phase 2,  $r(105) = -.29$ ,  $p = .002$ , jeweils zweiseitige Testung. Die

einzelne festgestellte signifikante Korrelation darf allerdings nicht kausal interpretiert werden. Da die Motivationsmaße *nach* der Steuerungsphase erhoben wurden, wäre (neben einer ursächlichen Drittvariablen) ebenso eine umgekehrte Kausalität denkbar. Beispielsweise hätte ein größerer subjektiv wahrgenommener Steuerungserfolg post hoc zu einer selbstwertdienlichen höheren Bewertung der Aufgabenanstrengung und der internalen Erfolgsattribution im Sinne eines *self-serving bias* (vgl. Tetlock & Levi, 1982) auf die eigene, als hoch eingeschätzte Anstrengung führen können.

Insgesamt gibt es in den Selbstauskunft-Daten zur Aufgabenmotivation somit keine konsistenten und stichhaltigen Hinweise darauf, dass kontextspezifische Motivationsgewinne oder Motivationsverluste in Experiment 1 eine Rolle für die Systemsteuerungsleistung in der Realgruppenbedingung spielten.

## Detaillierte Ablaufbeschreibung des Experimentes 2

Der experimentelle Ablauf war in allen wesentlichen Punkten identisch mit dem in Abbildung 6 dargestellten Ablauf des ersten Experimentes. Pro Erhebungstermin wurden bis zu 11 Probanden eingeladen, die von zwei oder drei Versuchsleitern betreut wurden. Alle Probanden wurden zu Beginn des Experimentes an separate Laptoparbeitsplätze in einem großen Versuchsraum gesetzt, begrüßt und über den Ablauf des Experimentes mit dem Ausfüllen der demographischen Angaben, der Systemexplorationsphase, der ersten individuellen Steuerungsphase, der Zuordnung zu einer der drei Bedingungen, der zweiten Steuerungsphase in der Individual-, Nominal-, oder Realbedingung und dem abschließenden Fragebogen informiert. Die standardisierte mündliche Instruktion entsprach weitestgehend der des ersten Experimentes. Es gab zwei experimentenspezifische Unterschiede in der Instruktion. Zum einen wurde über eine (im Vergleich zu Experiment 1 längere) durchschnittliche Versuchsdauer von 1,5 bis 2 Stunden, die mit einer (im Vergleich zu Experiment 1 höheren) Grundvergütung von 8 Euro einherging, informiert. Zum anderen erklärte die Versuchsleitung die Funktion des (im ersten Experiment nicht vorhandenen zusätzlichen) ausgedruckten Versuchsmaterials, das an jedem Laptoparbeitsplatz bereitgestellt wurde. Neben ihrem Laptop läge umgedreht ein Stapel Papier mit schriftlichen Zusatzinformationen zum System und mit Notizblättern. Im Verlauf des Experimentes würden sie aufgefordert, die Informationen und die Notizseiten zu nutzen. Die schriftlichen Informationen zum System ständen ihnen während des gesamten Experimentes zur Verfügung. Sie könnten sich während der Systemsteuerung Notizen zu ihren Überlegungen oder eventuell nötigen Nebenrechnungen auf dem dafür vorgesehenen Notizenblock machen.<sup>100</sup> Dieser Notizblock sei in Abschnitte entsprechend ihrer Steuerungseingaben unterteilt und Notizen sollten immer in den Notizbereich des Zeitabschnittes, für den sie gerade ein Urteil vornehmen würden, geschrieben werden. Die Verwendung eines Taschenrechners wurde untersagt. Nach der mündlichen Instruktion und eventueller Klärung von Verständnisfragen begannen die Probanden die computergestützte Bearbeitung des Experimentes. Der Ablauf war bis zur Zuordnung zu den Bedingungen

---

<sup>100</sup> Das systematische schriftliche Festhalten von Nebenrechnungen und Notizen über das System wurde angeboten, um der hohen Komplexität des Systems mit vielen Variablenbeziehungen Rechnung zu tragen und um die Daten bei Bedarf für eine explorative Analyse der Vorgehensweise der Probanden nutzen zu können. Diehl & Sterman (1995) hatten die Probanden in ihrer Untersuchung in ähnlicher Form auf Notebooks Notizen machen lassen.

nach der ersten Steuerungsphase für alle Probanden identisch und erfolgte im individuellen Arbeitstempo.

**Demographische Angaben.** Im Ablauf des Computerexperimentes gab es (abgesehen vom Inhalt der Steuerungsaufgabe) einen Unterschied zum ersten Experiment 1: Die Probanden machten anfangs nur demographische Angaben zu Geschlecht, Alter, Semesterzahl und Studienfach. Auf den NEO-FFI-Fragebogen wurde verzichtet, da die Auswertung des ersten Experimentes keine Relevanz der Persönlichkeitsfaktoren für die untersuchte Fragestellung ergeben hatte.

**Explorationsphase.** Vor der Explorationsphase lasen die Probanden am Bildschirm einen Einleitungstext zur Systemsteuerungsaufgabe, in dem ihnen als Gegenstand der Aufgabe des Experimentes die Steuerung bzw. das Management der Bereiche Produktion und Lagerhaltung eines Wirtschaftsunternehmens genannt wurde. Zunächst würden sie die Funktionsweise des Systems in einer Explorationsphase mit 20 Durchgängen testen können, um ein Gefühl für die Art der Steuerung zu bekommen. Nach einem Kurzüberblick über die weiteren Phasen des Experimentes und den leistungsabhängigen Bonus für die zweite Steuerungsphase wurden die Probanden gebeten die ausgedruckt vorliegenden grundlegenden Informationen zu den Systemeigenschaften zu lesen. Zur Explorationsphase erhielten die Probanden am Bildschirm den Hinweis, dass sie im Folgenden 20 Änderungen der Produktionsrate vornehmen sollten, um zu explorieren, wie das System genau funktionieren würde. Nach dieser Explorationsphase würden zwei Durchgänge folgen, in denen sie das System im Hinblick auf eine bestimmte Zielsetzung steuern sollten. Diese Zielvorgabe würde ihnen erst nach der Explorationsphase bekannt gegeben werden. Zusätzlich wurde den Probanden erklärt, worauf sie bei der Eingabe von Steuerungswerten in der Steuerungstabelle zu achten hatten und sie wurden darauf hingewiesen, dass sie Nebenrechnungen und Überlegungen zu Systemzusammenhängen und Funktionsweise des Systems auf dem ausgeteilten Notizblock, der in 20 Abschnitte entsprechend der 20 Steuerungseingaben unterteilt war, notieren durften. Anschließend gaben die Probanden sukzessive 20mal einen Wert für die „Änderung der Produktionsrate“ in die Steuerungstabelle ein. Nach jedem eingegebenen Wert drückten sie zur Bestätigung die Enter-Taste und klickten auf den Button „zur nächsten Einstellung“, woraufhin das System die resultierenden Systemwerte ausgab. Nach der zwanzigsten Eingabe erschien auf dem Monitor der Hinweis, dass die Testphase mit Durchgang 20 beendet sei, und dass die Probanden bitte „Weiter“ drücken sollten, um fortzufahren. Auf der folgenden Seite wurden



Fragen zu Wohlbefinden, Motivation, zum Vorgehen während der Explorationsphase und zu Annahmen der Probanden über das System gestellt, die im genauen Wortlaut denen des Experimentes 1 entsprachen-

**Steuerungsphase 1.** Durch einen Klick auf den „Weiter“-Button erschien die unter 5.3.1 genannte Instruktion für die individuelle Steuerung in der ersten Phase. Die Probanden wurden außerdem auf die Möglichkeit hingewiesen, den ausgeteilten Notizblock zu nutzen, der in 40 Abschnitte entsprechend der 40 Zeitabschnitte der Steuerung eingeteilt war. Die Probanden nahmen dann 40 Eingaben für die Steuerungsvariable „Änderung der Produktionsrate“ vor und verfolgten dabei das Ziel die kumulativen Gesamtkosten in der Ausgabetablelle minimal zu halten. Nach der ersten Steuerungsphase wurden die bereits nach der Explorationsphase gestellten Fragen zu Motivation und Annahmen über Systemzusammenhänge zum zweiten Mal in Bezug auf die Steuerungsphase 1 gestellt. Nach Klick auf den „Weiter“-Button erschien der Hinweis „Bitte rufen Sie jetzt die Versuchsleitung“. Die Versuchsleitung teilte die Probanden aufgrund ihrer individuellen Steuerungsleistung einer der drei Bedingungen zu.

**Zuteilung zu experimenteller Bedingung.** Alle Probandenlaptops waren über Wireless LAN mit einem Servercomputer vernetzt, auf dem die Versuchsleitung die individuellen kumulativen Gesamtkosten für die erste Steuerungsphase in einer Überblickstabelle einsehen konnte. Die Versuchsleitung ordnete drei Probanden einer Gruppe so zu, dass jeweils ein Gruppenmitglied mit einer relativ schlechten individuellen Steuerungsleistung, ein Mitglied mit einer mittelmäßigen Steuerungsleistung und ein Mitglied mit einer relativ guten Steuerungsleistung in der Gruppe vertreten waren. Für das Kriterium der Leistungsheterogenität gab es konkrete Leistungsgrenzen bezüglich der individuellen Ausgangsleistung. Das beste Gruppenmitglied sollte in der Steuerungsphase 1 einen Wert kleiner oder gleich 1 Million Euro erzielt haben, das mittlere Gruppenmitglied einen kumulativen Gesamtkostenwert zwischen 5 und 20 Millionen und das schlechteste Gruppenmitglied einen Wert zwischen 50 Millionen und 500 Milliarden (500 000 000 000) Euro. Probanden, deren Leistungswert sich keiner Dreiergruppe zuordnen ließ, wurden der Individualbedingung zugeteilt. Die gebildeten Dreiergruppen wurden zufällig entweder der Realbedingung oder der Nominalbedingung zugelost. Wie auch schon in Experiment 1 trugen die Versuchsleiter durch passwortgeschützte Eingabe an den Probandenlaptops die Bedingungszuordnung ein und gaben in der Nominal- und Realbedingung die Gruppenidentifikationsnummer an. Mit dieser Nummer konnte das Programm in der Nominalbedingung die individuellen Eingabewerte über den Serverrechner zu einem

Gruppenmittelwert verrechnen und diesen den drei jeweiligen Probandenlaptops zurückzumelden. In der Realbedingung diente die Gruppenidentifikationsnummer der späteren Sortierung der Probandendateien zur Auswertung auf Gruppenebene.

**Steuerungsphase 2 und abschließende Fragen in Einzelarbeit in der Individualbedingung.** Nach Einteilung in die Individualbedingung durch die Versuchsleitung blieben die Probanden an ihrem separaten Laptoparbeitsplatz sitzen und wurden am Monitor instruiert, nun über 40 weitere Zeittakte Produktion und Lager des Unternehmens zu steuern und die kumulativen Gesamtkosten möglichst minimal zu halten. Hierzu könnten sie 40mal einen Wert für die „Änderung der Produktionsrate“ eingeben. Analog zur ersten Steuerungsphase wurden die Probanden darauf hingewiesen, dass sie auch für die zweite Steuerungsphase einen Notizblock mit 40 Abschnitten zur Verfügung hätten, um sich Nebenrechnungen und Überlegungen zu notieren. Nach der Systemsteuerung über 40 Eingabeurteile im individuellen Tempo beantworteten die Probanden wieder Fragen zu Motivation und Annahmen über den Zusammenhang zwischen Änderung der Produktionsrate und den entstehenden Kosten während der Steuerungsphase. Im abschließenden Fragenteil wurde den Probanden ein Fragebogen zum Systemwissen mit 6 Multiple-Choice-Antworten dargeboten. Schließlich sollten die Probanden im offenen Antwortformat beschreiben, wie sie bei der Bearbeitung vorgegangen seien. Gleichwohl sollten sie ihre Kompetenz der zielgerichteten Systemsteuerung auf einer Skala von 1 = *sehr geringe Kompetenz* bis 7 = *sehr hohe Kompetenz* einschätzen. Außerdem wurde Vorerfahrung als Ausschlusskriterium mit der Frage erhoben: „Haben Sie schon einmal an einem Computereperiment teilgenommen, in dem es um die Steuerung eines Systems ging? Wenn ja, worum ging es genau?“. Die letzte Frage der Untersuchung bezog sich auf das vermutete Untersuchungsziel.

**Steuerungsphase 2 und abschließende Fragen in Einzelarbeit in der Nominalbedingung.** Die Probanden in der Nominalbedingung blieben an ihren individuellen Laptoparbeitsplätzen sitzen und hatten keine Kenntnis darüber, welche anderen zwei Probanden mit ihnen in eine Nominalgruppe eingeteilt worden waren. Sie erhielten auf dem Monitor die Instruktion, dass sie nun über 40 weitere Zeittakte Produktion und Lager des Unternehmens steuern sollen und dabei die Zielvorgabe hätten, die kumulativen Gesamtkosten möglichst minimal zu halten. Ihr Laptop sei in der zweiten Arbeitsphase über einen Server mit zwei anderen Versuchspersonencomputern verbunden. Jedes Gruppenmitglied dieser virtuellen Dreiergruppe nähme ohne Absprache mit den anderen Mitgliedern 40 individuelle Steuerungsurteile vor. Nach Eingabe ihres

individuellen Urteils erfolge keine Anzeige der Systemveränderungen. Es erschiene so lange „Bitte warten“, bis auch die anderen zwei Gruppenmitglieder ihre gewünschten Produktionsratenänderungswerte eingegeben hätten. Der Server würde dann die Einzelangaben in jedem Durchgang zu einem mittleren Gruppenurteil integrieren und diesen Gruppenwert in der zweiten Tabellenzeile zurückmelden. Es würde nur die aus dem gemittelten Gruppenurteil resultierende Systembeeinflussung zurückgemeldet werden. Sie könnten dann die individuelle Eingabe für den folgenden Durchgang vornehmen. Während der Systemsteuerung könnten die Probanden auf den ausgeteilten Notizblöcken Nebenrechnungen oder Überlegungen zu Systemzusammenhängen und zur Funktionsweise des Systems notieren. Nach der 40. Eingabe folge dann noch ein kurzer individueller Fragebogen.

Nach der Steuerungsphase wurden wie in allen Bedingungen Fragen zur Motivation und zu den getroffenen Annahmen über den Zusammenhang zwischen der Änderung der Produktionsrate und den entstehenden Kosten während der Steuerungsphase gestellt. Anschließend folgte der Fragebogen zum Systemwissen mit 6 Multiple-Choice-Antworten. Wie im Experiment 1 wurden die Nominalgruppenmitglieder in den abschließenden Fragen gefragt, wie wohl sie sich in der Gruppe gefühlt hätten und drittens wie wichtig es Ihnen gewesen wäre, dass die Einzelbeiträge aller drei Gruppenmitglieder in gleichem Maße in das Gruppenurteil einfließen. Mit den Angaben auf einer Skala von 1 = *sehr unwohl* bzw. *unwichtig* bis 7 = *sehr wohl* bzw. *sehr wichtig* zu den beiden letzten Fragen sollte Akzeptanz für die Nominalgruppensituation und die automatisierte Mittelung der Beiträge erfasst werden. Außerdem ermöglichte die Abfrage des Wohlbefindens einen Vergleich zur Realbedingung. Schließlich wurden die Probanden entsprechend der Formulierung in der Individualbedingung nach ihrer Vorerfahrung mit einer computergestützten Systemsteuerungsaufgabe und nach dem vermuteten Untersuchungsziel gefragt.

**Steuerungsphase 2 und abschließende Fragen in Einzelarbeit in der Realbedingung.** Entsprechend des beschriebenen Vorgehens im Experiment 1 wechselten jeweils drei Probanden, die derselben Realgruppe zugeordnet worden waren, mit ihren Laptops unter Anleitung eines Versuchsleiters in einen separaten Versuchsraum. Der Versuchsleiter wies den Gruppenmitgliedern Plätze an einem Gruppenarbeitstisch und entsprechend der Rangreihe ihrer Probandendurchlaufnummern im Experiment Schilder mit den Nummern 1, 2 oder 3 zu. Die Nummern dienten der anonymen Identifikation der Probanden, und die Probanden wurden gebeten sich die Zuordnung von Personen und Nummern zur Beantwortung späterer Fragen zu merken. Im Folgenden lasen die

Probanden auf ihren Laptopmonitoren die Instruktion zur zweiten Steuerungsphase. Demnach sollten sie über 40 weitere Zeittakte Produktion und Lager des Unternehmens steuern und hätten als Zielvorgabe, die kumulativen Gesamtkosten möglichst minimal zu halten. Für jeden Durchgang sollten sie zunächst individuell einen gewünschten Änderungswert der Produktionsrate überlegen und ihn in die erste Tabellenzeile eingeben. Währenddessen sollten sie sich nicht beraten oder beim Nachbarn auf den Bildschirm schauen. Nach Eingabe ihres individuellen Urteils erfolge *keine* Anzeige der Systemveränderungen. Wenn sie alle ihr individuelles Urteil eingegeben hätten, sollten sie ihre Vorschläge in der Form „Ich habe bei Durchgang 5 den Wert X eingegeben“ nennen und könnten anschließend ihre Ideen über die Funktionsweise des Systems und die optimale Eingabe diskutieren. Schließlich sollten sie sich auf einen Gruppenwert einigen, den sie alle drei in ihre Laptops in die zweite Zeile der Tabelle eingeben. Sie erhielten dann wie üblich eine Rückmeldung auf Basis der eingegebenen Gruppenurteile. Zusätzlich wurden die Probanden darauf hingewiesen, dass sie sich durch Nennen der Trialnummer vergewissern sollten, dass sich alle bei der Eingabe des gleichen Durchganges befänden und dass einmal getätigte Eingaben nicht mehr rückgängig gemacht werden konnten. Für das weitere Vorgehen bei Eingaben mit Tippfehlern wurde dasselbe Vorgehen wie in Experiment 1 vorgeschrieben. Außerdem wurden die Probanden wieder auf die Möglichkeit hingewiesen ihren persönlichen Notizblock mit 40 Abschnitten für die 40 Steuerungsdurchgänge der Phase 2 nutzen zu können. Die Versuchsleitung stehe für Fragen zur Verfügung, falls noch etwas unklar sein sollte. Ein Versuchsleiter blieb während der Gruppenarbeit im Raum, um die Einhaltung des vorgeschriebenen Vorgehens sicherzustellen und bei Missachtung korrigierend einzugreifen. Die abschließenden individuellen Fragen beantworteten die Realgruppenmitglieder wieder an getrennten Arbeitsplätzen ohne zu kommunizieren. Wie die Probanden in den anderen Bedingungen auch beantworteten sie Fragen zur Motivation während der zweiten Steuerungsphase, zu angenommenen Zusammenhängen zwischen der Änderung der Produktionsrate und den entstehenden Kosten sowie die sechs Multiple-Choice-Fragen des Fragebogens zum konkreten Systemwissen. Auch die zwei letzten Fragen des Abschlussfragebogens waren identisch zu den anderen zwei Bedingungen und bezogen sich auf die Vorerfahrung mit einer computergestützten Systemsteuerungsaufgabe und auf das vermutete Untersuchungsziel. Vor diesen zwei letzten Fragen erhielten die Probanden in der Realbedingung analog zum Ablauf im Experiment 1 jeweils eine Frage zum Wohlbefinden in der Gruppe, zur Wichtigkeit der gleichwertigen Berücksichtigung der Einzelbeiträge aller Gruppenmitglieder, zur wahrgenommenen Berücksichtigung ihrer eigenen Beiträge, zum

Vorgehen bei der Gruppenarbeit, zur wahrgenommenen relativen sowie zur absoluten Systemsteuerungskompetenz aller Gruppenmitglieder und zur subjektiven, prozentualen Gewichtung der Einzelbeiträge für das Gruppenurteil. Außerdem wurden die Probanden gefragt, welche Gruppenmitgliedsnummer sie selbst hatten, um eine korrekte Zuordnung und Auswertung der Kompetenzselbsteinschätzung sicherzustellen.

**Vergütung und Debriefing für alle Bedingungen.** Nachdem die Probanden den Abschlussfragebogen beantwortet hatten, wurde ihnen auf dem Computerbildschirm mitgeteilt, dass sie am Ende des Experimentes angekommen seien. Sie wurden gebeten, nicht mit anderen potenziellen Versuchspersonen über die Inhalte des Experimentes zu sprechen und erhielten Dank für Ihre Teilnahme. Die Versuchsleitung vergütete die Teilnahme mit einer Grundvergütung von 8 Euro oder alternativ mit einer Versuchspersonenstunde und zahlte zusätzlich den auf dem Serverrechner abgelesenen leistungsabhängigen Bonus aus, der zwischen 0 und 5 Euro pro Proband lag. Der Bonus betrug für Gesamtkostenwerte in der Steuerungsphase 2, die kleiner oder gleich 250000 Euro waren 5 Euro, für Gesamtkostenwerte, die größer 250000 und kleiner oder gleich 500000 Euro waren, 3 Euro und für Werte größer 500000 und kleiner oder gleich 1000000 1 Euro. Bei kumulierten Gesamtkosten über einer Million Euro betrug die Bonuszahlung 0

## Versuchsleistungsanweisung zu Experiment 2

### Ablaufplan für die Versuchsleitung und mündliche Instruktionen für die Versuchspersonen

#### 1. Checkliste vor dem Versuch: Vorbereitung

→ Benötigtes Material überprüfen, inkl. Schriftliche Instruktionen, Notizblöcke für die Explorationsphase und die Steuerungsphase 1 mit Versuchspersonennummer versehen und umgedreht neben den Laptops auslegen

→ Technik überprüfen

#### 2. Einleitende mündliche Instruktion (alle Bedingungen)

„Herzlich willkommen/Guten Morgen. Schön, dass Sie sich bereit erklärt haben, an der heutigen Untersuchung zum Thema „Systemsteuerungsaufgaben“ teilzunehmen. Uns interessiert, wie Personen individuell oder in der Gruppe diesen bestimmten Aufgabentyp bearbeiten. Die Untersuchung wird circa 1,5 bis 2 Stunden in Anspruch nehmen. Die Bearbeitungszeit hängt von der Versuchsbedingung und von Ihrem individuellen Arbeitstempo ab (Wenn nur Individualerhebung ca. 70 Minuten).

Das gesamte Experiment findet am Laptop statt. Es gliedert sich in vier Teile:

- Zunächst werden Sie gebeten einige demographische Angaben zu machen. Dann wird Ihnen das zu steuernde System schriftlich vorgestellt. Sie sind Manager in einem Unternehmen und für die Steuerung der Teilbereiche Produktion und Lagerhaltung verantwortlich. Anschließend können sich in 20 Übungsdurchgängen mit der Funktionsweise des Systems vertraut machen.
- Als drittes folgt eine individuelle Arbeitsphase, in der Sie über Durchgänge das System nach einer bestimmten Zielvorgabe steuern sollen. Nach dieser Phase wird Ihre persönliche Steuerungsleistung erfasst und Sie werden von der Versuchsleitung...
- ...in der vierten Arbeitsphase entweder einer individuellen oder einer Gruppenarbeitsbedingung zugewiesen:
  - [Individualbedingung:] In der individuellen Arbeitsbedingung steuern Sie erneut das System über Durchgänge alleine.  
[Je nach geplanten experimentellen Bedingungen nur die relevante Gruppenbedingung nennen]
  - [Nominalbedingung:] In einer Gruppenbedingung ist Ihr Computer über einen Server mit zwei anderen Versuchspersonencomputern verbunden. Jedes Gruppenmitglied nimmt räumlich getrennt ohne Absprache mit den anderen individuelle Steuerungsurteile vor, der Server integriert die Einzeleingaben pro Durchgang zu einem mittleren Gruppenurteil und meldet

- diesen Gruppenwert und die resultierende Systembeeinflussung nach jedem einzelnen der Durchgänge zurück.
- [Realbedingung:] In einer zweiten Gruppenbedingung werden Sie in Dreiergruppen auf die Nachbarräume aufgeteilt und sollen für jeden von Steuerungsdurchgängen zunächst ein individuelles Urteil treffen und dann in der Gruppe diskutieren und ein Gruppenurteil festlegen. Für den eingegebenen Gruppenwert erhalten Sie wie gewohnt eine Systemrückmeldung über die resultierenden Veränderungen der Systemvariablen.
  - Ganz am Ende folgt noch ein allgemeiner Fragebogen, den Sie individuell am PC ausfüllen.
  - Die Gruppen werden von der Versuchsleitung (mir) so zusammengestellt, dass Sie von der Leistung her möglichst gleich stark sind.
  - Für das Experiment gibt es zwei Versuchspersonenstunden/8 Euro Grundvergütung und einen leistungsabhängigen Bonus (0 bis 5 Euro), der davon abhängt, wie gut Sie das System nach der Zielvorgabe in der zweiten Arbeitsphase steuern. Um in der zweiten Phase gut abzuschneiden ist es sehr hilfreich, sich in der ersten individuellen Phase anzustrengen...

Neben Ihrem Laptop liegt umgedreht ein Stapel Papier mit schriftlichen Zusatzinformationen zum System und mit Notizblättern. Im Verlauf des Experimentes werden Sie aufgefordert, die Informationen und die Notizseiten zu nutzen.

Die schriftlichen Informationen zum System stehen Ihnen während des gesamten Experimentes zur Verfügung.

Sie können sich während der Systemsteuerung Notizen zu Ihren Überlegungen oder eventuell nötigen Nebenrechnungen auf dem dafür vorgesehenen Notizenblock machen. Dieser Notizblock ist in Abschnitte entsprechend Ihrer Steuerungseingaben unterteilt. Bitte schreiben Sie die Notizen immer in den Notizbereich des Zeitabschnittes, für den Sie gerade ein Urteil vornehmen. Die Verwendung eines Taschenrechners ist nicht erlaubt.

Falls Sie während des Versuches Fragen haben, bitte die Hand heben. Nach der individuellen Arbeitsphase erscheint auf dem Display „Bitte Versuchsleitung rufen“, auch dann bitte die Hand heben. Die Versuchsleitung/Ich teilt/teile Sie dann einer Bedingung für die zweite Steuerungsphase zu.

Gibt es jetzt noch Fragen? Dann können Sie beginnen. Viel Spaß!“

**3. Datentabelle auf Serverlaptop verfolgen, Leistungskennwerte der kumulierten Gesamtkosten nach Trial \*-40 ablesen → Gruppeneinteilung möglichst mit 3 unterschiedlichen Fähigkeitsniveaus, z.B. angrenzende Bonusbereiche 0, 1 und 3 (immer mindestens x% Abweichung zwischen allen paarweisen Vergleichen innerhalb einer Dreiergruppe, also z.B. ein Gruppenmitglied)**

Bereich 1: Kumulierte Kosten Phase 1 <= 100.000	Bonusbereich = 5
Bereich 2: Kumulierte Kosten Phase 1 <= 500.00	Bonusbereich = 3
Bereich 3: Kumulierte Kosten Phase 1 <= 1 Mio	Bonusbereich = 1
Bereich 4: Kumulierte Kosten Phase 1 > 1 Mio	Bonusbereich = 0

**Auf Laptops mit Passwort „test“ Bedingung (I, R, N) auswählen und Gruppennamen vergeben (realx; nominalx bzw. individualxxx (x ist in den Gruppenbedingungen die Gruppennummer, in der Individualbedingung die Versuchspersonennummer)**

**Gruppenmitglieder (Realgruppenbedingung) bitten Laptop und Maus mitzunehmen OHNE den Deckel herunterzuklappen (hierdurch geht die Serververbindung verloren), mit Laptops in kleinen Raum führen, Nummern eins bis drei auf Pappschildern zuweisen, wobei die niedrigste Versuchspersonendurchlaufnummer die 1 erhält (also z.B. Proband 170 die 1, 174 die 2, 177 die 3).**

**4. Notizblöcke für die Steuerungsphase 2 ausgeben (einen pro Person), VP-Nummer darauf notieren. Mündliche Instruktion:**

[Individualbedingung:] „Sie sollen nun über 40 weitere Zeittakte Produktion und Lager des Unternehmens steuern und die kumulativen Gesamtkosten möglichst minimal halten. Sie sind in die individuelle Arbeitsbedingung eingeteilt worden, in der Sie das System wie bisher allein steuern.“

[Nominalbedingung:] Sie sollen nun über 40 weitere Zeittakte Produktion und Lager des Unternehmens steuern und haben die Zielvorgabe, die kumulativen Gesamtkosten möglichst minimal zu halten. Ihr Laptop ist in der zweiten Arbeitsphase über einen Server mit zwei anderen Versuchspersonencomputern verbunden. Jedes Gruppenmitglied dieser virtuellen Dreiergruppe nimmt ohne Absprache mit den anderen 40 individuelle Steuerungsurteile vor. [Nur bei Nachfrage der Probanden, falls schriftliche Instruktionen unklar sind, weitere Zusatzklärung abgeben und ggf. genauer erläutern:] Nach Eingabe Ihres individuellen Urteils erfolgt *keine* Anzeige der Systemveränderungen. Es erscheint so lange „Bitte warten“, bis auch die anderen zwei Gruppenmitglieder ihre gewünschten Produktionsratenänderungswerte eingegeben haben. Der Server integriert die Einzeleingaben in jedem Durchgang zu einem mittleren Gruppenurteil und meldet diesen Gruppenwert in der zweiten Tabellenzeile zurück. Es wird nur die aus dem gemittelten



Gruppenurteil resultierende Systembeeinflussung zurückgemeldet. Sie können dann die individuelle Eingabe für den folgenden Durchgang vornehmen. Nach der 40. Eingabe folgt dann noch ein kurzer individueller Fragebogen.

[Realbedingung:] Sie sollen nun über 40 weitere Zeittakte Produktion und Lager des Unternehmens steuern und haben die Zielvorgabe, die kumulativen Gesamtkosten möglichst minimal zu halten. Für jeden Durchgang überlegen Sie zunächst individuell einen gewünschten Änderungswert der Produktionsrate und geben ihn in die erste Tabellenzeile ein. Währenddessen sollten Sie sich nicht beraten oder beim Nachbarn auf den Bildschirm schauen. Dann können Sie in der Gruppe diskutieren und einigen sich auf ein Gruppenurteil, das Sie in die zweite Tabellenzeile eingeben. Die Systemreaktion auf Ihr Gruppenurteil wird dann wie gewohnt angezeigt.“

[Nur bei Nachfrage der Probanden, falls schriftliche Instruktionen unklar sind, weitere Zusatzerklärung abgeben und ggf. genauer erläutern:] „Nach Eingabe Ihres individuellen Urteils erfolgt *keine* Anzeige der Systemveränderungen. Wenn Sie alle ihr individuelles Urteil eingegeben haben, tauschen Sie bitte Ihre Vorschläge in folgender Form aus: „Ich habe bei Durchgang 5 den Wert X eingegeben“. Wenn alle drei ihre individuellen Werte genannt haben, können Sie Ihre Ideen über die Funktionsweise des Systems und die optimale Eingabe diskutieren und sollen sich auf einen Gruppenwert einigen, den Sie alle drei in ihre Laptops in die zweite Zeile der Tabelle eingeben. Wenn Sie auf die Entertaste und dann auf „Bitte weiter“ klicken. Dann können Sie den Gruppenwert eingeben und erhalten wie üblich eine Temperaturrückmeldung auf Basis der eingegebenen Gruppenurteile. Bitte stellen Sie sicher, dass Sie sich immer bei der Eingabe des gleichen Durchganges befinden, indem Sie regelmäßig die Durchgangsnummer nennen.

Falls sich jemand in Ihrer Gruppe bei der Eingabe des Gruppenurteils vertippt, sollte er sich für die folgenden Durchgänge an den resultierenden Systemwerten der beiden anderen Gruppenmitglieder orientieren. Einmal getätigte und mit „Enter“ und „zur nächsten Einstellung“-Button bestätigte Eingaben können nicht mehr rückgängig gemacht werden.“

[Für alle Bedingungen:] „Dann viel Spaß, ich stehe für Fragen zur Verfügung, falls jetzt oder später noch etwas unklar sein sollte. Nach der 40. Eingabe folgt dann noch ein kurzer individueller Fragebogen.“

**5. Nach Beendigung kumulative Gesamtkosten der zweiten Steuerungsphase ablesen und entsprechenden Bonus zuordnen. Honorar auszahlen, Aufklärung über Untersuchungsziel (Lernprozesse in Gruppen, Vergleich der drei Bedingungen...siehe Debriefing-Bogen) und Verabschiedung.**

Bereich 1: Kumulierte Kosten Phase 1 <= 100.000	Bonusbereich = 5
Bereich 2: Kumulierte Kosten Phase 1 <= 500.00	Bonusbereich = 3
Bereich 3: Kumulierte Kosten Phase 1 <= 1 Mio	Bonusbereich = 1
Bereich 4: Kumulierte Kosten Phase 1 > 1 Mio	Bonusbereich = 0

**6. Datensicherung**

Daten sofort speichern (Passwort dafür ist ...).

Danach umbenennen der Datei für ein Format, mit dem sie automatisch in SPSS eingelesen werden kann. Dateiname vorne erweitern um Angabe zu Bedingung und – getrennt mit Unterstrichen- die Versuchspersonennummern; Beispiele:

- Real13\_059\_Laborlaptop4\_Dynamische Systeme\_20090701111234.xml
- Nominal01\_060\_Laborlaptop3\_Dynamische Systeme\_20090701112247.xml
- Individual\_099\_Laborlaptop4\_Dynamische Systeme\_20090708113710.xml

## **Debriefing Experiment Dynamische Systeme II**

In diesem Experiment untersuchen wir Gruppenleistung und Lernprozesse bei der Steuerung eines dynamischen Systems.

Bisherige Forschung zeigt, dass Einzelpersonen meistens nicht gut darin sind, Systeme nach bestimmten Zielvorgaben zu steuern bzw. Systemzustände zu kontrollieren. Sie unterliegen vielfältigen Urteilsfehlern und erzielen schlechtere Steuerungsergebnisse als nach einem optimalen Vorgehen oder auch durch Nichteingreifen in das System zu erreichen wären. Auch bei wiederholter Steuerung durchschauen Einzelpersonen Systemeigenschaften wie verzögerte Reaktion des Systems auf Eingriffe oder nichtlineare Beziehungen zwischen Systemvariablen meistens nicht.

Wir gehen der Frage nach, ob Gruppen bei der Steuerung dynamischer Systeme sogenannte „Prozessgewinne“ realisieren können, ob sie also eine bessere Steuerungsleistung erzielen als Individuen. Hierzu vergleichen wir die Leistung real interagierender Gruppen mit zwei weiteren Bedingungen. In der „Individualbedingung“ steuern je drei Probanden das System individuell und die gemittelten Leistungen werden mit der Gruppenleistung verglichen. In der „Nominalbedingung“ sind drei Personen, die individuelle Eingaben machen, über einen Serverrechner vernetzt. Aus ihren drei individuellen Urteilen in jedem Durchgang wird der Mittelwert gebildet, auf den das System dann reagiert. Diese Bedingung dient dazu zu erfassen, welche Vorteile allein durch die statistische Mittelwertbildung von drei Urteilen (ohne Interaktion in der Gruppe) entstehen. Eine darüber hinaus gehende bessere Realgruppenleistung kann dann auf Interaktionsprozesse in der Gruppe zurückgeführt werden.

Darüber hinaus untersuchen wir differenziert, welche Arten von Lernprozessen zur Realgruppenleistung beitragen. Zum einen kann es sein, dass die Gruppenmitglieder durch die Interaktion ihre individuelle Urteilsleistung verbessern (Man nennt dies Group-to-Individual-Transfer). Dies berechnen wir durch einen Vergleich der Individualleistung in den Phasen 1 und 2. Zum anderen kann es sein, dass Koordinationsgewinne auftreten und die Gruppe aus den drei individuellen Urteilsvorschlägen pro Durchgang ein noch besseres Steuerungsurteil abgibt als es der Mittelwert der drei Urteile wäre. Auch dies wird berechnet.

Über die offenen Fragen zu angenommenen Systemzusammenhängen erfassen wir zusätzlich das verbalisierbare Wissen und können auch hier einen Lernzuwachs über die drei Steuerungsphasen berechnen.

**Anhang I: Schriftliches Versuchsmaterial zu Experiment 2**

Grundlegende Probandeninformationen zum System

Gestaltung der Notizenseiten für die Explorationsphase

Gestaltung der Notizenseiten für die Steuerungsphase 1

Gestaltung der Notizenseiten für die Steuerungsphase 2

# GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

Georg-Elias-Müller-Institut für Psychologie  
Abteilung für Wirtschafts- und Sozialpsychologie



## Grundlegende Informationen zur Systemsteuerungsaufgabe

### „Produktions- und Lager - Management“

Sie sind Manager/in in einem Produktionsunternehmen und für die Bereiche Produktion und Lagerhaltung zuständig. Genereller Unternehmenszweck ist es, Güter zu produzieren und so die vorhandene Nachfrage zu befriedigen (also so viel zu produzieren, wie auch verkauft werden kann). In jeder Periode (Woche) der Systemsteuerung müssen Sie entscheiden, um wie viele Einheiten Sie die Produktionsrate erhöhen bzw. senken wollen. Die Produktionsrate stellt die Anzahl der in Auftrag gegebenen zu produzierenden Einheiten dar. Als Manager erhalten Sie für jede Urteilsperiode nach Festlegung der aktuellen Produktionsrate eine Rückmeldung über wichtige Variablenwerte des Produktions- und Warenlagersystems in tabellarischer Form (unten ohne Angabe konkreter Werte abgebildet). Die Ausgabe der Variablenwerte erfolgt immer gerundet auf ganze Stückzahlen bzw. auf ganze Euro-Werte.

Woche	1	2	3	4	5
Änderung der Produktionsrate	—	—	—	—	—
Aktuelle Produktionsrate					
<b>Produzierte Einheiten</b>					
Verkaufte Einheiten (abhängig)					
Verkaufte Einheiten (unabhängig)					
<b>Verkaufte Einheiten gesamt</b>					
Veränderung Lagerbestand					
<b>Lagerbestand</b>					
Produktionsanpassungskosten					
Lagerkosten					
<b>Gesamtkosten</b>					
<b>Kumulierte Gesamtkosten</b>					

**Bei der Steuerung am Computermonitor können Sie über Pfeiltasten unterhalb der Tabelle nach links und rechts scrollen und so den gesamten Bereich der bisher getätigten Steuerungsurteile und resultierenden Systemwerte überblicken.**

### **1. Informationen zur Produktion (Block 1 der Tabelle)**

Sie haben als Manager/in vollständige Kontrolle über die Produktion neuer Wareneinheiten. Zu Beginn Ihrer Systemsteuerung liegt die Produktionsrate bei 600 Einheiten. Über die *Änderung der Produktionsrate* können Sie die Stückzahl der in Auftrag gegebenen zu produzierenden Wareneinheiten erhöhen oder senken. Geben Sie z.B. „-50“ ein, wird die Produktion von 600 auf 550 neue Einheiten reduziert, geben Sie „+100“ ein, wird eine Steigerung der Produktionsrate von 600 auf 700 neue Wareneinheiten veranlasst. Die absolute Anzahl neu veranlasster Produktionseinheiten wird in der Zeile „*aktuelle Produktionsrate*“ angezeigt.

Achtung: Ein eingegebener negativer Wert für die Änderung der Produktionsrate kann maximal so groß sein, wie die zuvor vorhandene Produktionsrate. Liegt die veranlasste Produktionsrate in Woche 3 beispielsweise bei „1000“, kann die Änderung der Produktionsrate in Woche 4 maximal „-1000“ betragen. Sie können Ihre Produktion zwar auf 0 herunterfahren, aber natürlich ist es nicht möglich, dass Sie weniger als 0 Stück produzieren. Werden Eingaben vorgenommen, die außerhalb des möglichen negativen Wertebereiches liegen, korrigiert das System automatisch auf den niedrigsten möglichen Wert und zeigt neben dem Wert „(geändert)“ an.

In der Zeile „*Produzierte Einheiten*“ können Sie die Anzahl der in der aktuellen Urteilsperiode fertig gestellten Wareneinheiten ablesen.

### **2. Informationen zu den Verkäufen (Block 2 der Tabelle)**

Nicht erwartete Veränderungen der Verkaufszahlen verursachen eine unerwünschte positive oder negative Abweichung vom Lagerbestand 0. Es gibt zwei Arten von Warenverkäufen: Von Ihren eigenen Handlungen abhängige Verkäufe und von Ihren Handlungen unabhängige Verkaufszahlen.

**Die Anzahl abhängiger Verkäufe** wird von Ihren eigenen Produktionsentscheidungen beeinflusst. Die Produktion kann die Verkaufszahlen in zwei unterschiedliche Richtungen beeinflussen:

- Eine Erhöhung der Produktion kann zur **Steigerung** der Anzahl abhängiger Verkäufe führen: Wenn Sie die Produktionsrate steigern erhöhen Sie auch implizit die Anzahl Ihrer Arbeiter und zahlen mehr Löhne. Höhere Lohnauszahlungen kurbeln die Nachfrage nach den von Ihnen produzierten Waren an.
- Eine Erhöhung der Produktion kann zur **Senkung** der Anzahl abhängiger Verkäufe führen: Wenn Sie die Produktionsrate steigern, sehen auch Ihre Wettbewerber auf dem Markt die gesteigerten Absatzmöglichkeiten und steigern ihre Verkaufsanstrengungen: Erhöhte Wettbewerberaktivität durch Werbung etc. veranlasst einige Ihrer Kunden Bestellungen zu stornieren und bei der Konkurrenz zu kaufen. Dadurch verkaufen Sie selbst weniger Wareneinheiten.

Es tritt in nur einer der beiden Verkaufseffekte auf. Welcher Effekt dies ist, ist Ihnen im Vorfeld nicht bekannt.

**Die Anzahl unabhängiger Verkäufe** folgt einer zufälligen (d.h. für Sie nicht vorhersehbaren) Entwicklung. Die beste Vorgehensweise für Sie besteht darin anzunehmen, dass die unabhängigen Verkaufszahlen in der nächsten Periode identisch mit den aktuellen unabhängigen Verkaufszahlen sind. Tatsächlich können die unabhängigen Verkaufszahlen aber bis zu plus/minus

15 Einheiten von der vorherigen unabhängigen Absatzmenge abweichen. **Die Gesamtzahl an Verkäufen** beschreibt die Summe aller durch abhängige und durch unabhängige Verkäufe veräußerten Waren.

### 3. Informationen zum Lagerbestand (Block 3 der Tabelle)

Der Bestand des Warenlagers, das Sie verwalten, wird durch Warenverkäufe reduziert und durch produzierte Einheiten erhöht. Sie können Ihr Verkaufsvolumen nur zum Teil vorhersagen, aber Sie haben vollständige Kontrolle über die Produktion neuer Wareneinheiten.

**Die Veränderung des Lagerbestandes** zeigt an, ob sich in der aktuellen Steuerungsperiode die Stückzahl der Waren im Lager erhöht hat (positive Werte) oder gesunken ist (negative Werte).

**Der Lagerbestand-Wert** zeigt den absoluten Bestand an Wareneinheiten in der aktuellen Steuerungsperiode an. Zu Beginn Ihrer Steuerung liegt der Lagerbestand bei 0 Einheiten.

### 4. Informationen zur Kostenstruktur (Block 4 der Tabelle)

Kosten können zwei Entstehungsursachen haben: Veränderungen der Produktionsmenge und Nichtübereinstimmung von Lagerbestand und Nachfrage. Um die Kosten nach bestimmten Zielvorgaben kontrollieren zu können, müssen Sie diese zwei Kostenkomponenten genau verstehen:

**Veränderungen der Produktionsrate:** Sie verursachen **Produktionsanpassungskosten**. Dies sind Folgekosten für die Entlassung von Mitarbeitern bei Senkung der Produktionsrate bzw. Kosten für die Einstellung und Einarbeitung neuer Mitarbeiter bei Erhöhung der Produktionsrate.

**Nichtübereinstimmung von Lagerbestand und Nachfrage:** Ist der tatsächliche Lagerbestand größer als 0 entstehen **Lagerkosten** für die fertigen Waren. Wenn der tatsächliche Lagerbestand kleiner als 0 ist, sind keine Waren verfügbar, wenn der Kunde sie kaufen möchte (d.h. „-10“ bedeutet, dass 10 Einheiten bestellt worden sind, die momentan nicht geliefert werden können). Das verursacht Irritation beim Kunden und bedeutet Verluste für das Unternehmen (wenn ungeduldige Kunden z.B. zur Konkurrenz abwandern).

Ihnen ist im Vorfeld nicht bekannt, ob die gleichen Kosten anfallen, wenn sie ihre Produktion um 10 Einheiten erhöhen oder um 10 Einheiten reduzieren. Auch wissen Sie nicht, ob ein bestimmter Lagerbestand größer 0 (z.B. 20 Einheiten) genauso große Kosten verursacht wie ein gleich großer negativer Bestand, bei dem Sie 20 Kunden nicht beliefern können.

Außerdem wissen Sie im Vorfeld nicht, ob größere Veränderungen der Produktionsmenge und größere Abweichungen des Lagerbestands von Null überproportional hohe Kosten verursachen oder ob einer Veränderungseinheit der Produktionsmenge bzw. einer Abweichungseinheit vom optimalen Lagerbestand von 0 immer ein identischer Wert in Euro zuzuordnen ist. In welchem Verhältnis die Kosten für die Änderung der Produktionsrate zu den Kosten durch Nichtübereinstimmung von Lagerbestand und Nachfrage stehen, ist Ihnen im Vorfeld ebenfalls nicht genau bekannt.

Die **Gesamtkosten** einer Steuerungsperiode ergeben sich aus der Summe von Produktionsanpassungskosten und Lagerkosten. Die **kumulierten Gesamtkosten** in der letzten Tabellenzeile beschreiben die bis zur aktuellen Steuerungsperiode aufsummierten Gesamtkosten.

**Dieses Informationsblatt können Sie auch weiterhin während der Systemexploration und –steuerung einsehen. Bitte klicken Sie auf den „Weiter“-Button auf dem Bildschirm, um mit der Systemexploration zu beginnen.**

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

Georg-Elias-Müller-Institut für Psychologie  
Abteilung für Wirtschafts- und Sozialpsychologie



Code: \_\_\_\_\_

**Notizen zum Steuerungsurteil Woche 1 (Explorationsphase)**

**Notizen zum Steuerungsurteil Woche 2 (Explorationsphase)**



Code: \_\_\_\_\_

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

Georg-Elias-Müller-Institut für Psychologie  
Abteilung für Wirtschafts- und Sozialpsychologie



**Notizen zum Steuerungsurteil Woche 1 (Steuerungsphase 1)**

**Notizen zum Steuerungsurteil Woche 2 (Steuerungsphase 1)**

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

Georg-Elias-Müller-Institut für Psychologie  
Abteilung für Wirtschafts- und Sozialpsychologie



Code: \_\_\_\_\_

**Notizen zum Steuerungsurteil Woche 1 (Steuerungsphase 2)**

**Notizen zum Steuerungsurteil Woche 2 (Steuerungsphase 2)**

## Kodieranweisung zu den Wissensfragen in Experiment 2

Kodiert werden die Antworten im Datensatz auf Individualebene, pro Zeile in SPSS sind nacheinander also die Antworten von einer Versuchspersonen zu finden.

Für jede Versuchsperson wird separat kodiert.

In der SPSS-Datei sind bereits Kodiervariablen zwischen den Antwortspalten eingefügt, bei 0 vs. 1 Kodierungen bitte auch die Null eintragen.

### Je einmal pro Zeile zu kodierende Items in chronologischer Reihenfolge:

- **Nach Übungsphase mit 20 Durchgängen ohne Zielvorgabe**
  - **vorgeh\_uephase**  
→Wie sind Sie in der ersten Explorationsphase des Systems mit den 20 Zeittakten vorgegangen?
  - **annahme\_1**  
→Wie lauteten Ihre konkreten Annahmen über den Zusammenhang von Stellradeinstellung und Kühltemperatur? Falls sich Ihre Annahmen im Verlauf der Steuerung geändert haben, ordnen Sie sie bitte zeitlich so, dass die aktuell gültige Annahme an letzter Stelle steht. Wenn Sie weniger als 3 Annahmen haben, schreiben Sie bitte "x" in die nicht benötigten Textfelder Wenn Sie mehr als drei Annahmen haben, schreiben Sie diese bitte mit Nummerierung (4., 5. usw.) in das dritte Textfeld. Bitte nennen Sie alle Annahmen, die Sie während der letzten Steuerungsphase hatten, (auch wenn diese während der Explorationsphase oder der ersten Steuerungsphase schon vorhanden waren!).
  - **annahme\_2**
  - **annahme\_3**
  
- **Nach der ersten Aufgabenphase mit 40 Durchgängen**
  - **annahme\_1\_phase1**
  - **annahme\_2\_phase1**
  - **annahme\_3\_phase1**
  
- **Nach der zweiten Aufgabenphase mit 40 Durchgängen (individuell oder in der Gruppe)**
  - **annahme\_1\_phase2**
  - **annahme\_2\_phase2**
  - **annahme\_3\_phase2**
  - **ziel** → Was glauben Sie war das Ziel dieser Untersuchung?
  - **Materialok** → Fanden Sie Teile der Aufgabenstellung und/oder der Fragen unverständlich?
  - **vorgehen** → Wie sind Sie bei der Bearbeitung vorgegangen? Haben Sie zum Beispiel spezifische Erfahrungen aus der ersten individuellen Arbeitsphase genutzt?

## 5. Kodierung Vorgehen Übungsphase/Phase 2

- (7) = keine Angabe/x/Null
- (8) = Beobachtungsphase (mehrere Durchgänge geschaut, wie das System reagiert; Eigendynamik zu erkennen versucht; über mehrere Durchgänge konstanten Wert eingegeben und abgewartet)
- (9) = Systematische Eingriffsvariation (absichtlich hohen bzw. niedrigen Wert eingegeben und geschaut, wie das System reagiert; mehrmals konstanten Wert eingegeben etc.)
- (10) = Zeitverzögerte Systemreaktion erkannt und genutzt
- (11) = Vorgehen nach Gefühl/intuitiv
- (12) = falsche Hypothesen/System nicht erkannt, daher kein planmäßiges Vorgehen
- (13) = Ja, Erfahrungen genutzt, aber keine kodierbare Spezifikation (keine genauere Angabe)

## 6. Kodierung der Systemannahmen (für jedes Annahmefeld in der Übungsphase, Phase 1 und Phase 2)

### 2.a Richtigkeit der einzelnen Annahmen pro Feld in zwei Spalten kodieren

(1 für vorhanden, 0 für nicht vorhanden kodieren)

- $r\_annahme_{1/2/3\_m1\_ue/ph1/ph2}$  **Annahme richtig**  
→ (1 für zutreffend, 0 für nicht zutreffend)
- $f\_annahme_{1/2/3\_m1\_ue/ph1/ph2}$  **Annahme falsch**  
(1 für zutreffend, 0 für nicht zutreffend)  
→ Wenn in einem Annahmefeld eine richtige und eine falsche Annahme stehen mit „1“ in beiden Feldern kodieren
- $annahme_{1/2/3\_m1\_ue/ph1/ph2\_dich}$   
→ 0, wenn  $r\_annahme = 0$  und  $f\_annahme = 0$   
→ 1, wenn  $r\_annahme = 1$  und  $f\_annahme = 0$   
→ 2, wenn  $r\_annahme = 0$  und  $f\_annahme = 1$   
→ 3, wenn  $r\_annahme = 1$  und  $f\_annahme = 1$

**2.b Inhaltskategorien kodieren** (nach jeder Annahme 17 Spalten mit Indizes 0 bis 17; 1 für vorhanden, 0 für nicht vorhanden in jeder Spalte kodieren; grüne Kodierkategorien sind inhaltlich richtig, rote inhaltlich falsch):

**Itembeispiele, wenn mit „1“ kodiert:**

- $annahme_{1\_1\_m1\_ue}$  : Zusammenhang richtig (1) bei Annahme 1 des ersten Gruppenmitglieds in der Übungsphase
- $annahme_{1\_3\_m1\_ph1}$  : ...richtige Aussage (3) bei Annahme 1 des ersten Gruppenmitglieds in der Phase 1
- $annahme_{2\_7\_m1\_ph2}$  : Zufall angenommen (7) bei Annahme 2 des zweiten Gruppenmitglieds in der Phase 2

- **(1) Zusammenhang Produktionskosten richtig:** „Größere Veränderungen der Produktionsrate führen zu höheren Produktionsanpassungskosten“/ "Veränderungen um +/- 4 Einheiten verursachen gleich hohe Anpassungskosten" (nur Richtung wichtig bzw. gleiches Kostenverhältnis von Veränderungswerten nach oben und unten; nicht Funktion angesprochen, dazu siehe Kategorie (3))  
**(Erhöhung und Senkung haben vergleichbare Konsequenzen:** „ Eine Erhöhung der Produktionsrate um x (z.B. 10) Einheiten führte zu *genauso großen* Produktionsanpassungskosten wie eine Senkung der Produktionsrate um x (z.B. 10) Einheiten)
- **(2) Zeitverzögerte Reaktion des Systems richtig** (z.B. „System reagiert erst später/2 Wochen später“)
- **(3) Exponentialfunktion Produktionskosten richtig:**(z.B. „Große Änderungen der Produktionsmenge führten zu überproportional hohen Kosten.“/ „überproportional hohe Kosten bei größeren Veränderungsschritten“)
- **(4) Zusammenhang Produktionskosten falsch** (z.B sehr kleine und sehr große Anpassungen der Produktionsrate verursachen höhere Kosten als Anpassungen im mittleren Bereich; eine Erhöhung um x Einheiten führt zu höheren Kosten als eine Senkung um x Einheiten)
- **(5) Zeitverzögerte Reaktion des Systems falsch** (z.B. „System reagiert sofort auf Eingaben“)
- **(6) Magisches Zahlendenken** (z.B. „gerade Zahlen = Kostensteigerung, ungerade Zahlen = Kosteneinsparung“)
- **(7) Zufall/kein schlüssiges System** (z.B. „willkürlich?“)
- **(8) Zusammenhang Lagerbestand richtig** (z.B. „ Ein bestimmter Lagerbestand x (z.B. 20 Einheiten) größer 0 verursachte *genauso große* Kosten wie ein gleich großer negativer Bestand von  $-x$  Einheiten, bei dem Sie (z.B. 20) Kunden nicht beliefern konnten.“)
- **(9) Exponentialfunktion Lagerkosten richtig** (z.B. „Große Abweichungen des Lagerbestandes vom Optimum 0 (im positiven oder negativen Bereich) führten zu überproportional hohen Kosten.“)
- **(10) Kostenverhältnis Produktion und Lagerkosten richtig (z.B.** „Eine Abweichung des Lagerbestandes um x Einheiten (z.B. 10 Einheiten) führte zu *halb so großen Kosten* wie eine Änderung der Produktionsrate um x (z.B. 10) Einheiten.“)

- **(11) Positive Gain bei abhängigen Verkäufen richtig** („Eine Erhöhung der Produktion führte zur **Steigerung** der Anzahl abhängiger Verkäufe.“)
- **(12) Zusammenhang Lagerbestand falsch** (z.B. „Ein bestimmter Lagerbestand  $x$  (z.B. 20 Einheiten) größer 0 verursachte *höhere/niedrigere* Kosten als ein gleich großer negativer Bestand von  $-x$  Einheiten, bei dem Sie (z.B. 20) Kunden nicht beliefern konnten.
- **(13) Kostenverhältnis Produktion und Lagerkosten falsch (z.B.** „Eine Abweichung des Lagerbestandes um  $x$  Einheiten (z.B. 10 Einheiten) führte zu *genauso großen* Kosten wie eine Änderung der Produktionsrate um  $x$  (z.B. 10) Einheiten.“ ODER „Abweichungen des Lagerbestandes beeinflussten die Gesamtkosten stärker als Veränderungen in der Produktion.“)
- **(14) Positive Gain bei abhängigen Verkäufen falsch** („Eine Erhöhung der Produktion führte zur **Senkung** der Anzahl abhängiger Verkäufe.“)
- **(0) optimale Strategie zur Kostenminimierung richtig:** (Produktionsrate möglichst in kleinen Schritten anpassen, evtl. Begründung: Anpassungskosten fallen stärker ins Gewicht als Lagerkosten, Zeitverzögerung einkalkulieren, Lagerbestand sollte möglichst gering gehalten werden)
- **(00) Grundannahme: Das System ist nicht stabil/Kosten steigen stetig auch bei Nichteingreifen)**

→ Falls eine Aussage unklar/nicht schlüssig für den Kodierer ist, unter diffussem Systemverständnis mit (7) kodieren!

→ Wenn pro Annahmefeld mehrere Annahmen aufgeschrieben werden, sollten beide kodiert werden, also z.B. für annahme1\_ue (1) und (2) mit Einserkodierung in den Spalten annahme1\_1\_m1\_ue und annahme1\_2\_m1\_ue, entsprechend bei Kodierung richtig/falsch auch als zwei Annahmen berücksichtigen

→ Wenn Annahmen innerhalb einer Arbeitsphase (z.B. Annahmen 1 bis 3 der Phase 1) mehrfach genannt werden, nur beim ersten Mal kodieren).

→ Kontrolle vollständiger Kodierung: Anzahl der gesamten Annahmen gleich Anzahl der Inhaltskodierungen pro Block von drei Annahmen!

## 7. Kodierung Ziel

(3) = falsche/keine Zielangabe

(4) = richtig: Vergleich individueller und Gruppenleistung

(5) = richtig: Leistung/Lernfortschritt bei der Systemsteuerung

**8. Kodierung Materialok****„Waren Teile der Instruktionen/Fragen unverständlich?“**

(4) = nein (nö, x, -, ...)

(5) = ja, sonstige Kritik an Verständlichkeit der Frageitems oder Instruktionen  
(Materialkritik am „neuen“ Fragebogenmaterial der offenen Fragen)

(6) = Aufgabe zu schwer, Zusammenhänge des Systems unverständlich etc. (bezieht sich auf Bewältigungsfähigkeit, nicht auf Verständlichkeit der Aufgabenstellung)

**→ Auffällige Angaben mit Probanden-ID gesondert notieren in letzter Spalte**

### Ergänzende Analysen zu Lerneffekten auf individueller Ebene in Experiment 2

**Lerneffekte auf individueller Ebene.** Zur detaillierteren Analyse des individuellen Leistungszuwachses wurden post-hoc Einzel-*t*-Tests gegen Null für das jeweils beste, mittlere und schlechteste Mitglied für jede Bedingung berechnet. Die Ergebnisse sind in der Übersicht in Tabelle 15 aufgeführt.

Tabelle 15: *Differenzmaßwerte und Ergebnisse separater Einstichproben-t-Tests gegen Null für den individuellen Leistungszuwachs in der Steuerungsphase 2 des Experiments 2*

Bedingung	<i>M</i> ( <i>SD</i> )	<i>t</i> (34)	<i>p</i>	<i>Cohen's d</i>
Real				
bestes Mitglied	-.35 (1.85)	-1.12	.271	-0.38
mittleres Mitglied	3.39 (2.63)	7.64	< .001	0.19
schlechtestes Mitglied	6.77 (3.07)	13.06	< .001	2.21
gesamt	3.27 (2.28)	8.48	< .001	1.43
Nominal				
bestes Mitglied	-1.92 (2.11)	-5.38	< .001	-0.91
mittleres Mitglied	1.97(2.17)	5.38	< .001	0.91
schlechtestes Mitglied	5.90 (2.70)	12.92	< .001	2.19
gesamt	1.98 (2.04)	5.74	< .001	0.97
Individual				
bestes Mitglied	.12 (1.03)	.71	.485	0.12
mittleres Mitglied	.94 (2.49)	2.22	.033	0.38
schlechtestes Mitglied	2.17 (3.36)	3.81	.001	0.65
gesamt	1.08 (1.36)	4.67	< .001	0.79



---

In der Individualbedingung wiesen das in Steuerungsphase 1 mittlere und schlechteste Gruppenmitglied Übungseffekte auf, die mit  $p = .033$  bzw.  $p = .001$  signifikant von Null verschieden waren. Das anfangs beste Gruppenmitglied zeigte hingegen keinen I-I Transfer,  $p = .485$ . Die mittleren sowie die schlechtesten Mitglieder der Real- und der Nominalbedingung manifestierten einen noch stärkeren individuellen Leistungszuwachs (alle  $p < .001$ ) als in der Individualbedingung. Auffällig ist, dass die besten Mitglieder in der Real- und der Nominalbedingung im Durchschnitt schlechter in der Phase 2 als in der Phase 1 abschnitten. In der Realbedingung verschlechterte sich das beste Gruppenmitglied nur deskriptiv,  $p = .271$ . In der Nominalbedingung verschlechterte sich das beste Mitglied hingegen signifikant mit  $p < .001$ . Dies änderte nichts daran, dass über alle Gruppenmitglieder hinweg in jeder der drei Bedingungen signifikante individuelle Leistungssteigerungen von der ersten zur zweiten Steuerungsphase vorlagen, alle  $p < .001$ .

### Explorative Analyse: Zusammenhang von motivationalen Variablen und Steuerungsleistung in Experiment 2

Wie in den vorherigen Experimenten schätzten die Probanden individuell ihr Wohlbefinden und ihre Aufgabenmotivation auf einer siebenstufigen Skala ein, wobei 1 eine sehr geringe und 7 eine sehr hohe Ausprägung bedeutete. Post hoc wurde für jede Angabe der Gruppenmittelwert pro Phase berechnet.

Die durchgeführte MANOVA mit einem 3 (*Steuerungsphase*: Explorationsphase vs. Phase 1 vs. Phase 2) x 3 (*Bedingung*: Real vs. Nominal vs. Individual) Design wurde für den Phasenverlauf der abhängigen Variablen Wohlbefinden signifikant,  $F(2, 102) = 106.91$ ,  $p < .001$ ,  $\eta^2_p = .51$ . Abbildung 22 illustriert, dass das Wohlbefinden von der zweiten zur dritten Phase in allen Bedingungen anstieg. Es lag kein signifikanter Bedingungseffekt vor,  $F(2, 102) = 1.46$ ,  $p = .238$  und  $\eta^2_p = .03$ . Jedoch trat eine signifikante Interaktion zwischen der Steuerungsphase und dem Faktor Bedingung auf,  $F(2, 102) = 11.94$ ,  $p < .001$ ,  $\eta^2_p = .19$ . Sowohl ein diesbezüglicher linearer, als auch ein quadratischer Innersubjektcontrast waren signifikant, beide  $F(2, 102) > 56.23$ , beide  $p < .004$ , beide  $\eta^2_p > .35$ . Das subjektive Wohlbefinden blieb in allen Bedingungen von der Explorationsphase zur ersten Steuerungsphase auf gleichem Niveau und stieg dann in der zweiten Steuerungsphase an. In der Realbedingung gaben die Gruppenmitglieder einen deutlich größeren Anstieg des Wohlbefindens in der Gruppenphase an als die Probanden in den anderen beiden Bedingungen. Wie im ersten Experiment lag eine signifikante Korrelation zwischen der gruppenbasierten Steuerungsleistung in Phase 2 und dem mittleren Wohlbefinden auf Gruppenebene vor: Die Beurteilung des Wohlbefindens nach relativ erfolgreicher Systemsteuerung fiel höher aus,  $r(105) = -.33$ ,  $p = .001$ , zweiseitige Testung. Eine spezifische Analyse des Wohlbefindens in Phase 2 zeigte mindestens einen signifikanten Unterschied zwischen den drei Bedingungen auf,  $F(2,102) = 13.35$ ,  $p < .001$ ,  $\eta^2_p = .21$ . Post hoc Einzelvergleiche zeigten, dass die Realgruppenmitglieder sich signifikant wohler fühlten ( $M = 5.37$ ,  $SD = .83$ ) als die Probanden in der Nominal- ( $M = 4.37$ ,  $SD = .87$ ) und der Individualbedingung ( $M = 4.44$ ,  $SD = 1.01$ ), beide  $p < .001$ . Letztere zwei Bedingungen unterschieden sich nicht im Wohlbefinden,  $p = 1$ .

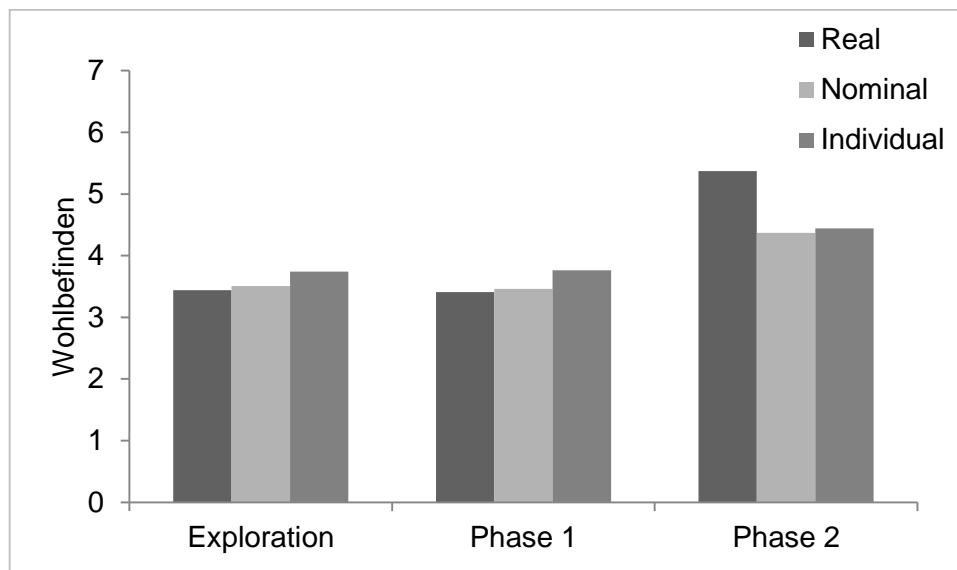


Abbildung 22. Subjektives Wohlbefinden in der Explorationsphase, der Steuerungsphase 1 und der Steuerungsphase 2 getrennt nach Bedingungen in Experiment 2.

In einer zweiten MANOVA wurde die subjektiv wahrgenommene Wichtigkeit, das dynamische System zu verstehen (als Motivationsindikator) betrachtet. Die pro Gruppe bzw. Nominalgruppe gemittelten Probandeneinschätzungen gingen als abhängige Variable in eine 3 (Innersubjektfaktor *Steuerungsphase*: Explorationsphase vs. Phase 1 vs. Phase 2) x 3 (*Bedingung*: Real vs. Nominal vs. Individual) MANOVA ein. Es lag hierbei kein signifikanter Haupteffekt für die Steuerungsphase vor,  $F(1.74, 177.36) = 0.20$ ,  $p = .792$ ,  $\eta^2_p = .00$ . Die Bedingungen unterschieden sich nicht generell in der Verständnismotivation,  $F(2, 102) = 2.51$ ,  $p = .086$ ,  $\eta^2_p = .05$ . Allerdings wurde die Interaktion zwischen von Steuerungsphase und der Bedingung signifikant,  $F(3.48, 177.36) = 4.85$ ,  $p = .002$ ,  $\eta^2_p = .09$ . In der Individual- und der Nominalbedingung reduzierte sich die Verständnismotivation von der ersten zur zweiten Steuerungsphase. In der Realbedingung war die Motivation hingegen nach der Explorationsphase relativ gering ausgeprägt und steigerte sich dann über die erste bis zur zweiten Phase. Ein separater Bedingungsvergleich für die Phase 2 konnte keinen signifikanten Unterschied der subjektiven Verständnis Wichtigkeit nachweisen,  $F(2, 102) = 0.80$ ,  $p = .450$ ,  $\eta^2_p = .02$ .

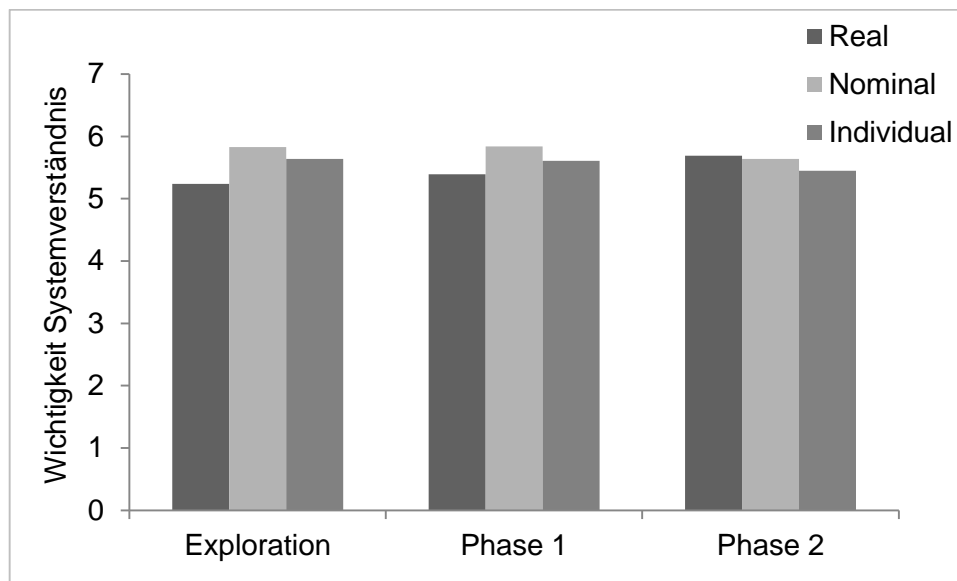


Abbildung 23. Subjektive Wichtigkeit, das System zu verstehen in der Explorationsphase, der Steuerungsphase 1 und der Steuerungsphase 2 getrennt nach Bedingungen in Experiment 2.

Als dritte Variable wurde die subjektive Anstrengung bzw. Mühe, die sich die Probanden bei der Systemsteuerung in der der Explorations- und den zwei Steuerungsphasen gegeben hatten, ausgewertet. Die mittleren Gruppenratings gingen als abhängige Variable in eine 3 (Innersubjektfaktor *Steuerungsphase*: Explorationsphase vs. Phase 1 vs. Phase 2) x 3 (*Bedingung*: Real vs. Nominal vs. Individual) MANOVA ein. Der Effekt für die Steuerungsphase war signifikant,  $F(1.31, 133.86) = 13.13, p < .001, \eta^2_p = .11$ . In Abbildung 23 sind die Werte von der Explorationsphase bis zur zweiten Steuerungsphase nach Bedingungen getrennt dargestellt. Für die Steuerungsphase lag mit  $F(2, 102) = 0.55, p = .580, \eta^2_p = .01$ , kein signifikanter Haupteffekt für den Faktor Bedingung vor. Der Interaktionseffekt von Phase x Bedingung wurde ebenfalls nicht signifikant,  $F(2.62, 133.68) = 4.85, p = .406, \eta^2_p = .02$ . Die Verständnismotivation nahm in allen drei Bedingungen von der ersten zur zweiten Steuerungsphase leicht zu (vgl. Abb. 24). Eine separate Betrachtung der subjektiven Mühe in Phase 2 zeigte keinen Bedingungsunterschied auf,  $F(2, 102) = 2.67, p = .074, \eta^2_p = .05$ .

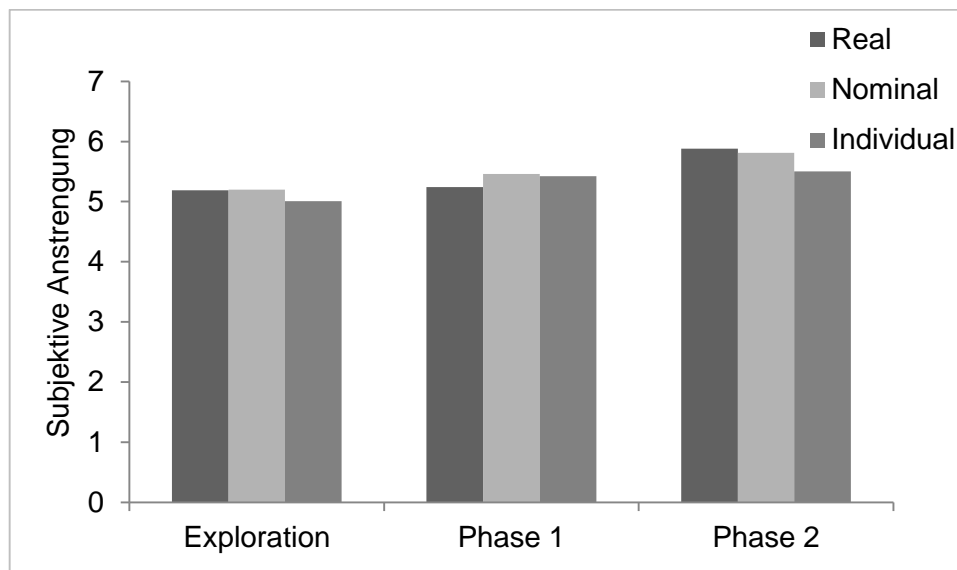


Abbildung 24. Subjektive Anstrengung bei der Systemexploration bzw. -steuerung in der Explorationsphase, der Steuerungsphase 1 und der Steuerungsphase 2 getrennt nach Bedingungen in Experiment 2.

Schließlich wurde für jeden der Motivationsindikatoren betrachtet, ob eine signifikante Korrelationen mit dem Steuerungoutput der log2-transformierten Gesamtkosten der Steuerungsphase 2 bestand. Es bestand kein Zusammenhang zwischen der Verstehensmotivation und der Steuerungsleistung,  $r(105) = -.11$ ,  $p = .262$ . Das Ausmaß der subjektiven Anstrengung bzw. gegebenen Mühe korrelierte hingegen signifikant mit der Leistung, und  $r(105) = -.25$ ,  $p = .010$  (jeweils zweiseitige Testung). Für die Interpretation des Zusammenhangs gelten die bei der Auswertung der Motivationsmaße in Experiment 1 getroffenen Aussagen: Eine Kausalinterpretation des nach der Steuerungsleistung erhobenen subjektiven Motivationsmaßes als ursächlich für eine höhere Leistung ist nicht möglich.

## Versuchsleistungsanweisung zu Experiment 3

### Ablaufplan für die Versuchsleitung/Instruktionen für die Versuchspersonen

#### 1. Checkliste vor dem Versuch:

- Material vorhanden?
- Blatt „Gruppenzuordnung und Bonusdokumentation“ vorhanden
- Technikcheck, insbesondere Kameraaufbau

#### 2. Einleitende mündliche Instruktion (alle Bedingungen)

„Herzlich willkommen/Guten Morgen. Schön, dass Sie sich bereit erklärt haben, an der heutigen Untersuchung zum Thema „Systemsteuerungsaufgaben“ teilzunehmen. Uns interessiert, wie Personen individuell oder in der Gruppe diesen bestimmten Aufgabentyp bearbeiten. Die Untersuchung wird circa 1,5 – 2,5 Stunden in Anspruch nehmen. Die Bearbeitungszeit hängt von der Versuchsbedingung und von Ihrem individuellen Arbeitstempo ab (Wenn nur Individualerhebung ca. 90 Minuten).

Das gesamte Experiment findet am Laptop statt. Es gliedert sich in fünf Teile:

- Zunächst werden Sie gebeten einige **demographische Angaben** zu machen und einige Fragen zu Ihrer Person zu beantworten
- Dann wird Ihnen die Funktionsweise des zu steuernden Systems schriftlich erklärt und Sie können sich in **20 Übungsdurchgängen** mit der Funktionsweise vertraut machen. Es geht um die Steuerung eines Kühlsystems in einem Atomkraftwerk. Dabei handelt es sich um eine Optimierungsaufgabe und die Ziele bestehen darin erstens so schnell wie möglich einen Optimalwert zu erreichen und ihn dann möglichst gut beizubehalten. Es sollen also die Temperaturabweichungen vom Optimalwert über alle Steuerungsdurchgänge möglichst gering gehalten werden. Nach den Übungsdurchgängen folgen drei Steuerungsphasen:
- Zuerst findet eine **individuelle Arbeitsphase statt**, in der Sie über 60 Durchgänge das System nach der bestimmten Zielvorgabe steuern sollen. Nach dieser Phase wird Ihre persönliche Steuerungsleistung erfasst und Sie werden von der Versuchsleitung...
- ...in der **zweiten Steuerungsphase** entweder einer individuellen oder einer Gruppenarbeitsbedingung zugewiesen:
  - [Individualbedingung:] In der individuellen Arbeitsbedingung steuern Sie erneut das System über 60 Durchgänge alleine.  
[Je nach geplanten experimentellen Bedingungen nur die relevante Gruppenbedingung nennen]
  - [Nominalbedingung:] In einer Gruppenbedingung ist Ihr Computer über einen Server mit zwei anderen Versuchspersonencomputern verbunden.

Jedes Gruppenmitglied nimmt räumlich getrennt ohne Absprache mit den anderen 60 individuelle Steuerungsurteile vor, der Server integriert die Einzeleingaben pro Durchgang zu einem mittleren Gruppenurteil und meldet diesen Gruppenwert und die resultierende Systembeeinflussung nach jedem einzelnen der 60 Durchgänge zurück. Dann erfolgt die Eingabe für den nächsten Durchgang usw.

- [Realbedingung:] In einer zweiten Gruppenbedingung werden Sie in Dreiergruppen auf die Nachbarräume aufgeteilt und sollen für jeden von 60 Steuerungsdurchgängen zunächst ein individuelles Urteil treffen und dann in der Gruppe diskutieren und ein Gruppenurteil festlegen. Für den eingegebenen Gruppenwert erhalten Sie wie gewohnt eine Systemrückmeldung über die resultierende Temperatur.
- Nach der zweiten Steuerungsphase durchlaufen alle Teilnehmer noch eine **dritte individuelle Steuerungsphase**, in der Sie das System wieder allein steuern.
- Ganz am Ende folgt noch ein **allgemeiner Fragebogen**, den Sie individuell am PC ausfüllen.
  
- Die Gruppen in der zweiten Phase werden von der Versuchsleitung (mir) so zusammengestellt, dass Sie von der Leistung her möglichst gleich stark sind.
- Für das Experiment gibt es zwei Versuchspersonenstunden/10 Euro Grundvergütung und für die Steuerungsphasen zwei und drei **jeweils** einen leistungsabhängigen Bonus (0 bis 5 Euro), der davon abhängt, wie gut Sie das System nach der Zielvorgabe in der zweiten Arbeitsphase steuern. Insgesamt können Sie also 20 Euro in diesem Experiment verdienen. Zur Bonusberechnung wird die mittlere prozentuale Abweichung der 60 erreichten Temperaturwerte von der anzustrebenden Optimaltemperatur bestimmt. Um in der zweiten und dritten Phase gut abzuschneiden ist es sehr hilfreich, sich in der ersten individuellen Phase anzustrengen...

Bitte achten Sie darauf, dass Sie mit der Maus immer ganz nach unten scrollen und dann auf den „Weiter-Button“ klicken. Sollte der Button einmal nicht funktionieren, haben Sie wahrscheinlich vergessen, eine Frage zu beantworten/ein Häkchen zu setzen. Wenn Sie offene Fragen nicht beantworten möchten, setzen Sie bitte ein „x“ in das Antwortfeld.

Wenn Sie während des Versuches Fragen haben, bitte die Hand heben. Nach der individuellen Arbeitsphase erscheint auf dem Display „Bitte Versuchsleitung rufen“, auch dann bitte die Hand heben...

Gibt es jetzt noch Fragen? Dann können Sie beginnen. Viel Spaß!

**3. Datentabelle auf Serverlaptop verfolgen. Wenn abgeschlossen, in Officetabelle kopieren, Leistungskennwerte ablesen und auf dem Blatt „Gruppenzuordnung und Bonuszahlung“ MAPE sowie Gruppenzuordnung der Versuchspersonen eintragen.**

- Gruppeneinteilung soll mit 3 unterschiedlichen Fähigkeitsniveaus, z.B. angrenzende Bonusbereiche 0, 1 und 3 erfolgen. Es sollen mindestens 10% Abweichung zwischen allen paarweisen Vergleichen innerhalb einer Dreiergruppe vorliegen, also z.B. ein Gruppenmitglied mit einem MAPE von 2.10, eins mit einem MAPE von .99 und eins mit einem MAPE von .50 (Kriterium „Startkompetenz der Gruppen“)
- Differenz zwischen dem schlechtesten und dem besten Gruppenmitglied soll zwischen .30 (besser .50) und 2.00 liegen (Kriterium „Heterogenität der Gruppe“)

MAPE Phase1 $\leq .35$	(entspricht Bonus = 5)
MAPE Phase 1 $> .35$ and $x \leq .81$	(entspricht Bonus = 3)
MAPE Phase 1 $> .81$ and $x \leq 1.24$	(entspricht Bonus = 1)
MAPE Phase 1 $> 1.24$	(entspricht Bonus = 0)

Auf Laptops mit Passwort „test“ Bedingung (I, R, N) auswählen und Gruppennamen vergeben (realx; nominalx bzw. individualxxx ( x ist in den Gruppenbedingungen die Gruppennummer, in der Individualbedingung die Versuchspersonennummer)

Gruppenmitglieder (Realgruppenbedingung) bitten Laptop und Maus mitzunehmen OHNE den Deckel herunterzuklappen (hierdurch geht die Serververbindung verloren), mit Laptops in kleinen Raum führen, Nummern eins bis drei auf Pappschildern zuweisen, wobei die niedrigste Versuchspersonendurchlaufnummer die 1 erhält (also z.B. Proband 170 die 1, 174 die 2, 177 die 3).

Gruppenidentitätsübersicht filmen, anschließend Kamera auf Gruppe richten.

#### 4. Mündliche Instruktion:

[Individualbedingung:] „Sie sollen nun über 60 weitere Zeittakte das Kühlsystem steuern und möglichst den Optimalwert von 4 Grad Celsius erreichen. Sie sind in die individuelle Arbeitsbedingung eingeteilt wurden, in der Sie das System wie bisher allein steuern.“

[Nominalbedingung:] Sie sollen nun über 60 weitere Zeittakte das Kühlsystem steuern und möglichst den Optimalwert von 4 Grad Celsius erreichen. Ihr Laptop ist in der zweiten Arbeitsphase über einen Server mit zwei anderen Versuchspersonencomputern verbunden. Jedes Gruppenmitglied dieser virtuellen Dreiergruppe nimmt ohne Absprache mit den anderen 60 individuelle Steuerungsurteile vor, der Server integriert die Einzeleingaben zu einem mittleren Gruppenurteil und meldet diesen Gruppenwert und die resultierende Systembeeinflussung nach jedem einzelnen der 60 Durchgänge zurück. Nach



Ihrer ersten Eingabe erscheint so lange „Bitte warten“, bis auch die anderen zwei Gruppenmitglieder ihre Stellradwerte eingegeben haben. Dann wird Ihnen der gemittelte Stellradwert und die daraus resultierende Temperatur zurückgemeldet und Sie können die Eingabe für den folgenden Durchgang durchführen. Nach der 60. Eingabe folgt dann noch ein kurzer individueller Fragebogen.

[Realbedingung:] Sie sollen nun über 60 weitere Zeittakte das Kühlsystem steuern und möglichst den Optimalwert von 4 Grad Celsius erreichen. Für jeden Durchgang überlegen Sie zunächst individuell einen Stellradwert und geben ihn ein. Währenddessen sollten Sie sich nicht beraten oder beim Nachbarn auf den Bildschirm schauen. Wenn Sie alle ihr individuelles Urteil eingegeben haben, tauschen Sie bitte Ihre Vorschläge in folgender Form aus: „Ich habe bei Durchgang 5 den Wert X eingegeben“. Wenn alle drei ihre individuellen Werte genannt haben, können Sie Ihre Ideen über die Funktionsweise des Systems und die optimale Eingabe diskutieren und sollen sich auf einen Gruppenwert einigen, den Sie alle drei in ihre Laptops eingeben. Der individuelle Eingabewert verschwindet, wenn Sie auf „OK“ klicken. Dann können Sie den Gruppenwert eingeben und erhalten wie üblich eine Temperaturrückmeldung auf Basis der eingegebenen Gruppenurteile. Bitte stellen Sie sicher, dass Sie sich immer bei der Eingabe des gleichen Durchganges befinden, indem Sie regelmäßig die Durchgangsnummer nennen.

Falls sich einer in Ihrer Gruppe bei der Eingabe des Gruppenurteils vertippt, sollte er sich für die folgenden Durchgänge an den resultierenden Temperaturen der beiden anderen Gruppenmitglieder orientieren. Einmal getätigte Eingaben können nicht mehr rückgängig gemacht werden.

Ich werde die Gruppenarbeit auf **Video** aufzeichnen. [Gruppenidentitätsblatt ausfüllen und auf Video aufnehmen, bevor die Interaktion beginnt] Die Videoaufzeichnung dient nur Auswertungszwecken und wird nicht an Dritte weiter gegeben. Dadurch soll sichergestellt werden, dass nur solche Gruppen in die Auswertung einbezogen werden, die sich an die Instruktionen gehalten haben. Bitte unterhalten Sie sich ganz normal und beachten Sie die Kamera nicht.

Dann viel Spaß, ich stehe für Fragen zur Verfügung, falls jetzt oder später noch etwas unklar sein sollte. Nach dem 60. Steuerungsurteil beende ich die Videoaufzeichnung und Sie gehen zur Steuerung der dritten Phase wieder in den individuellen Arbeitsraum.

**5. Nach Beendigung der zweiten Phase Daten der zweiten Steuerungsphase in Officetabelle kopieren, Leistungskennwerte/Bonuszahlung ablesen und auf dem Blatt „Gruppenzuordnung und Bonuszahlung“ eintragen.**

**Nach Beendigung des Experimentes Daten der dritten Steuerungsphase in Officetabelle kopieren, Leistungskennwerte/Bonuszahlung ablesen und auf dem Blatt „Gruppenzuordnung und Bonuszahlung“ eintragen.**

**Gesamtvergütung ausrechnen und Honorar auszahlen, Aufklärung über Untersuchungsziel (Lernprozesse in Gruppen, Vergleich der drei Bedingungen...siehe Debriefing-Bogen) und Verabschiedung.**

## **6. Datensicherung**

Daten sofort speichern (Passwort dafür ist ...).

Danach umbenennen der Datei für ein Format, mit dem sie automatisch in SPSS eingelesen werden kann. Dateiname vorne erweitern um Angabe zu Bedingung und – getrennt mit Unterstrichen- die Versuchspersonennummern; Beispiele:

- Real13\_059\_Laborlaptop4\_Dynamische Systeme\_20090701111234.xml
- Nominal01\_060\_Laborlaptop3\_Dynamische Systeme\_20090701112247.xml
- Individual\_099\_Laborlaptop4\_Dynamische Systeme\_20090708113710.xml



---

## Experiment Dynamische Systeme 3

### Gruppenidentitätsübersicht

**Datum:** \_\_\_\_\_

**Uhrzeit Beginn:** \_\_\_\_\_

**Versuchsleiter(in):** \_\_\_\_\_

**Gruppennummer:** \_\_\_\_\_

**Versuchspersonennummer Gruppenmitglied 1:** \_\_\_\_\_

**Versuchspersonennummer Gruppenmitglied 2:** \_\_\_\_\_

**Versuchspersonennummer Gruppenmitglied 3:** \_\_\_\_\_

**Dauer der Gruppenarbeitsphase (bis Urteil 60):** \_\_\_\_\_ min.

**Bemerkungen/Auffälligkeiten:**

Anmerkung:

Gruppenmitglieder sind von links nach rechts um den Tisch angeordnet.

### Debriefing Experiment Dynamische Systeme III

In diesem Experiment untersuchen wir Gruppenleistung und Lernprozesse bei der Steuerung eines dynamischen Systems.

Bisherige Forschung zeigt, dass Einzelpersonen meistens nicht gut darin sind, Systeme nach bestimmten Zielvorgaben zu steuern bzw. Systemzustände zu kontrollieren. Sie unterliegen vielfältigen Urteilsfehlern und erzielen schlechtere Steuerungsergebnisse als nach einem optimalen Vorgehen oder auch durch Nichteingreifen in das System zu erreichen wären. Auch bei wiederholter Steuerung durchschauen Einzelpersonen Systemeigenschaften wie verzögerte Reaktion des Systems auf Eingriffe oder nichtlineare Beziehungen zwischen Systemvariablen meistens nicht.

Wir gehen der Frage nach, ob Gruppen bei der Steuerung dynamischer Systeme sogenannte „**Prozessgewinne**“ realisieren können, ob sie also eine bessere Steuerungsleistung erzielen als Individuen. Hierzu vergleichen wir die Leistung real interagierender Gruppen mit zwei weiteren Bedingungen. In der „Individualbedingung“ steuern je drei Probanden das System individuell und die gemittelten Leistungen werden mit der Gruppenleistung verglichen. In der „Nominalbedingung“ sind drei Personen, die individuelle Eingaben machen, über einen Serverrechner vernetzt. Aus ihren drei individuellen Urteilen in jedem Durchgang wird der Mittelwert gebildet, auf den das System dann reagiert. Diese Bedingung dient dazu zu erfassen, welche Vorteile allein durch die statistische Mittelwertbildung von drei Urteilen (ohne Interaktion in der Gruppe) entstehen. Eine darüber hinaus gehende bessere Realgruppenleistung kann dann auf Interaktionsprozesse in der Gruppe zurückgeführt werden.

Darüber hinaus untersuchen wir differenziert, welche Arten von Lernprozessen zur Realgruppenleistung beitragen. Zum einen kann es sein, dass die Gruppenmitglieder durch die Interaktion ihre individuelle Urteilsleistung verbessern (Man nennt dies Group-to-Individual Transfer). Dies berechnen wir durch einen Vergleich der Individualleistung in den Phasen 1 und 2. Zum anderen kann es sein, dass Koordinationsgewinne auftreten und die Gruppe aus den drei individuellen Urteilsvorschlägen pro Durchgang ein noch besseres Steuerungsurteil abgibt als es der Mittelwert der drei Urteile wäre. Auch dies wird berechnet.

Über die offenen Fragen zu angenommenen Systemzusammenhängen erfassen wir zusätzlich das verbalisierbare Wissen und können auch hier einen Lernzuwachs über die drei Steuerungsphasen berechnen.

Die dritte individuelle Steuerungsphase dient dazu festzustellen, ob der Group-to-Individual Transfer stabil ist und auch ohne alternierende Gruppeninteraktion zwischen den individuellen Urteilen bestehen bleibt.

### **Explorative Analyse: Zusammenhang von motivationalen Variablen und Steuerungsleistung in Experiment 3 und Experiment übergreifende Analysen**

Die Probanden wurden an drei Zeitpunkten während des Experimentes gebeten individuell ihr Wohlbefinden und ihre Aufgabenmotivation auf einer siebenstufigen Skala einzuschätzen, wobei 1 eine sehr geringe und 7 eine sehr hohe Ausprägung indizierte. Die Aufgabenmotivation wurde über die Wichtigkeit, das System zu verstehen bzw. über die gegebene Mühe bei der vorangehenden Aufgabenbearbeitung erfasst. Post hoc wurde für jede der drei Angaben pro Gruppe der Mittelwert der Mitgliederangaben berechnet.

Eine MANOVA mit einem 3 (*Steuerungsphase*: Explorationsphase vs. Phase 1 vs. Phase 2) x 3 (*Bedingung*: Real vs. Nominal vs. Individual) Design zeigte für die abhängige Variable Wohlbefinden einen signifikanten Unterschied im Phasenverlauf,  $F(2, 204) = 106.91$ ,  $p < .001$ ,  $\eta^2_p = .52$ . Dieser wurde durch einen linearen sowie einen quadratischen Innersubjektkontrast qualifiziert, beide  $p < .001$ , beide  $F > 56.23$  und  $\eta^2_p > .35$ . Abbildung 25 illustriert, dass das Wohlbefinden von der zweiten zur dritten Phase in allen Bedingungen anstieg. Es gab keinen signifikanten Bedingungseffekt,  $F(2, 102) = 1.46$ ,  $p = .238$  und  $\eta^2_p = .03$ , aber eine signifikante Interaktion zwischen Phase und Bedingung mit  $F(4, 204) = 11.94$ ,  $p < .001$ ,  $\eta^2_p = .19$ . Sowohl ein diesbezüglicher linearer, als auch ein quadratischer Innersubjektkontrast waren signifikant, beide  $F(2, 102) > 6.23$ , beide  $p < .002$ , beide  $\eta^2_p > .19$ . Das subjektive Wohlbefinden in allen Bedingungen blieb von der Explorationsphase zur ersten Steuerungsphase stabil und steigerte sich dann in der zweiten Steuerungsphase. In der Realbedingung verzeichneten die Gruppenmitglieder einen deutlich größeren Anstieg des Wohlbefindens in der Gruppenphase als in den anderen beiden Bedingungen. Wie in Experiment 1 auch, bestand eine signifikante Korrelation zwischen der gruppenbasierten Steuerungsleistung in Phase 2 und dem mittleren Wohlbefinden auf Gruppenebene. Probanden beurteilen ihr Wohlbefinden nach relativ erfolgreichem Systemmanagement höher,  $r(105) = -.33$ ,  $p = .001$ , zweiseitige Testung. Eine kausale Interpretation des Zusammenhangs ist nicht möglich. Eine separate Analyse für das Wohlbefinden in Steuerungsphase 2 wurde mit  $F(2, 102) = 14.03$ ,  $p < .001$  und  $\eta^2_p = .22$  signifikant.

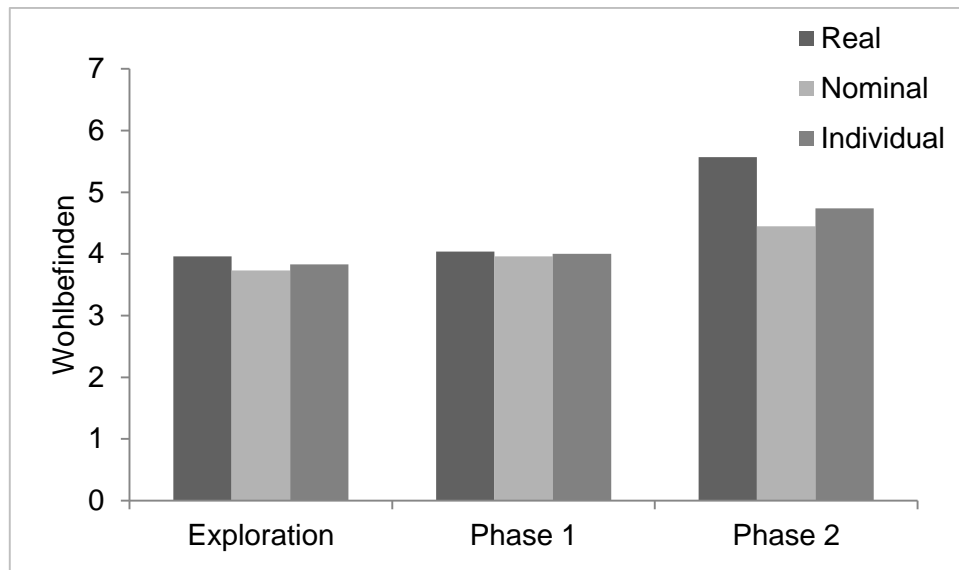


Abbildung 25. Subjektives Wohlbefinden in der Explorationsphase, der Steuerungsphase 1 und der Steuerungsphase 2 getrennt nach Bedingungen in Experiment 3.

Als nächstes wurde die subjektiv wahrgenommene Wichtigkeit, das dynamische System zu verstehen (als ein Indikator für die Aufgabenmotivation) betrachtet. Die auf Gruppenebene gemittelten Ratings der Probanden gingen als abhängige Variable in eine 3 (Innersubjektfaktor *Steuerungsphase*: Explorationsphase vs. Phase 1 vs. Phase 2) x 3 (*Bedingung*: Real vs. Nominal vs. Individual) MANOVA ein. Der Haupteffekt für die Steuerungsphase war nicht signifikant,  $F(2, 204) = .19, p = .823, \eta^2_p = .09$ . Die Bedingungen unterschieden sich auch nicht, in der Verständnismotivation,  $F(2, 102) = 2.51, p = .086, \eta^2_p = .05$ . Die Interaktion zwischen von Steuerungsphase x Bedingung war jedoch signifikant,  $F(4, 204) = 4.85, p < .001, \eta^2_p = .09$ . Während die Wichtigkeit, ein gutes Systemverständnis zu erlangen genauso wie in Experiment 1 in der Individual- und der Nominalbedingung von der ersten zur zweiten Steuerungsphase abnahm, zeigte sich für die Realbedingung ein anderer Verlauf. Die Motivationsverläufe für das Erlangen von Systemverständnis sind in Abbildung 26 zu sehen. In der Realbedingung war die Motivation nach der Explorationsphase relativ gering ausgeprägt und steigerte sich dann über die erste bis zur zweiten Phase. Die gesonderte Analyse der Phase 2 zeigte einen signifikanten Bedingungseffekt,  $F(2, 102) = 3.69, p = .028$  und  $\eta^2_p = .07$ . Dieser beruhte auf einem signifikanten Unterschied zwischen der Realbedingung ( $M = 6.14, SD = .66$ ) und der Nominalbedingung ( $M = 5.63, SD = .76$ ) mit  $p = .028$ . Die beiden anderen Bedingungsvergleiche waren nicht signifikant, beide  $p > .486$ . In der Individualbedingung lag die mittlere subjektiv wahrgenommene Wichtigkeit das System zu verstehen zwischen der Ausprägung in der Real- und der Nominalbedingung bei  $.590 (SD = .93)$ .

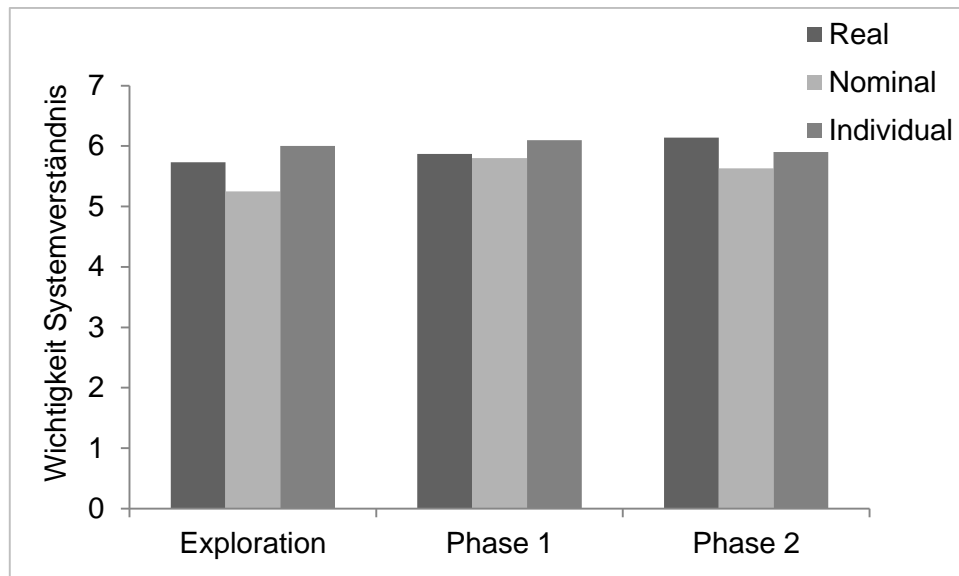


Abbildung 26. Subjektive Wichtigkeit, das System zu verstehen in der Explorationsphase, der Steuerungsphase 1 und der Steuerungsphase 2 getrennt nach Bedingungen in Experiment 3.

Als letzte Variable wurde die subjektive Mühe, die sich die Probanden bei der Systemsteuerung in der jeweils vorangehenden Phase gegeben hatten, ausgewertet. Die mittleren Gruppenratings gingen ebenfalls als abhängige Variable in eine **3** (Innersubjektfaktor *Steuerungsphase*: Explorationsphase vs. Phase 1 vs. Phase 2) x **3** (*Bedingung*: Real vs. Nominal vs. Individual) MANOVA ein. Der Effekt für die Steuerungsphase war auch hier signifikant,  $F(2, 204) = 13.13, p < .001, \eta^2_p = .11$ , und mit einem signifikanten linearen Innersubjektcontrast verbunden,  $F(1, 102) = 18.58, p < .001, \eta^2_p = .49$ . In Abbildung 27 ist die entsprechende kontinuierliche Motivationssteigerung von der Explorationsphase bis zur zweiten Steuerungsphase ersichtlich. Die separate Analyse der Steuerungsphase 2 zeigte eine vergleichbare subjektiv gegebene Mühe zwischen den Bedingungen auf,  $F(2, 102) = 1.37, p = .258, \eta^2_p = .03$ .

Anschließend wurde für jeden der Motivationsindikatoren betrachtet, ob eine signifikante Korrelationen mit dem Steuerungoutput der MAPE-Werte in der Steuerungsphase 2 bestand. Dies war weder für die Verstehensmotivation noch für die Anstrengungsmotivation bzw. gegebene Mühe der Fall,  $r(105) = -.15, p = .126$  und  $r(105) = -.12, p = .221$ , jeweils zweiseitige Testung. Somit besteht auch für das dritte Experiment kein Anlass davon auszugehen, dass die Motivation der steuernden Probanden einen bedeutsamen Anteil an den Leistungsunterschieden zwischen den Experimentalbedingungen hatte.



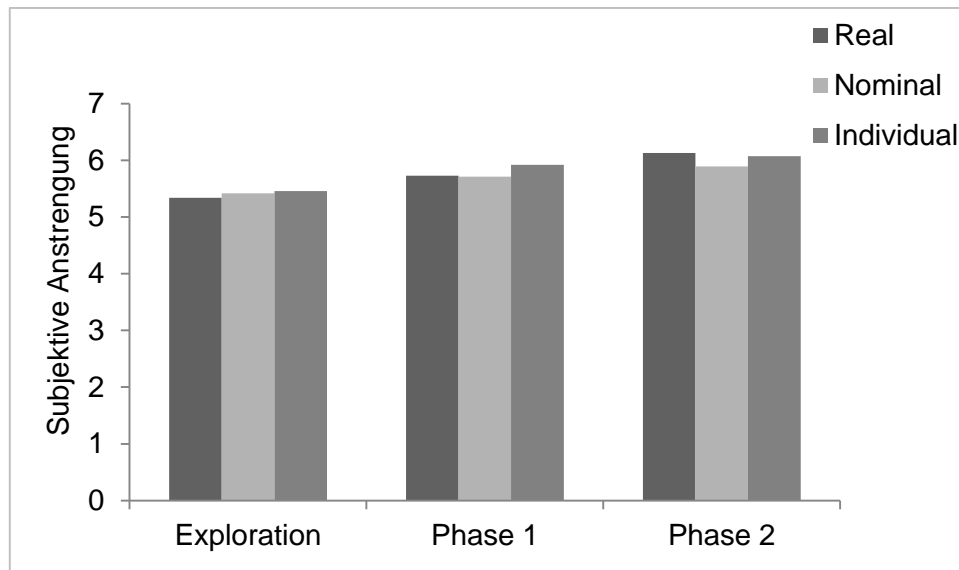


Abbildung 27. Subjektive Anstrengung in der Explorationsphase, der Steuerungsphase 1 und der Steuerungsphase 2 getrennt nach Bedingungen in Experiment 3.

In Kapitel 6.4.1 der Arbeit wurde berichtet, dass das generelle Leistungslevel der Systemsteuerung des Kühlsystemszenarios in Experiment 3 höher lag als in Experiment 1. Eine mögliche Erklärung für dieses höhere Leistungsniveau könnte eine gesteigerte Leistungsmotivation der Probanden gewesen sein. Im Folgenden wurden daher die subjektiven Angaben der Probanden zu Ihrer Motivation im Fragebogen nach den Steuerungsphasen ausgewertet. In drei getrennten  $t$ -Tests wurde die subjektiv aufgewandte Mühe der Probanden (eingeschätzt auf einer Skala von 1 *sehr wenig Mühe* bis 7 *sehr viel Mühe*) jeweils für die Explorationsphase, die Steuerungsphase 1 und die Steuerungsphase 2 berechnet. Der Vergleich für die subjektive Bearbeitungsmotivation in der Explorationsphase war mit  $t(208) = -1.54$ ,  $p = .125$  und  $d = 0.21$  nicht signifikant. Die berichtete Mühe in Experiment 3 ( $M = 5.41$ ,  $SD = .68$ ) fiel lediglich deskriptiv höher aus als in Experiment 1 ( $M = 5.27$ ,  $SD = .63$ ). Der Vergleich für die Aufgabenmotivation in der ersten individuellen Steuerungsphase war hingegen mit  $t(208) = -2.92$ ,  $p = .004$  und  $d = 0.41$  signifikant. Die Probanden in Experiment 3 gaben sich nach eigener Angabe mehr Mühe ( $M = 5.79$ ,  $SD = .59$ ) als die Probanden in Experiment 1 ( $M = 5.51$ ,  $SD = .77$ ). Die subjektive Aufgabenmotivation in der Steuerungsphase 2 über alle Experimentalbedingungen hinweg unterschied sich noch deutlicher zwischen den Experimenten 1 ( $M = 5.69$ ,  $SD = .75$ ) und 3 ( $M = 6.03$ ,  $SD = .65$ ),  $t(208) = -3.54$ ,  $p < .001$  und  $d = 0.49$ . Eine Erklärung für die bedingungsübergreifend höhere Aufgabenmotivation im dritten Experiment könnte die größere Aufgabenrelevanz aufgrund der drei Steuerungsphasen und damit verbundener höherer Maximalbonuszahlungen für die Probanden gewesen sein. Eine ergänzende bzw. alternative Erklärung bot der Erhebungszeitpunkt des Experimentes 3 der zeitlich nahe an der Atomkraftwerkkatastrophe in

Fukushima, Japan, lag. Das Ereignis könnte die subjektive Bedeutsamkeit der Aufgabe für die Probanden und damit auch die Motivation erhöht haben. Für die Interpretation der Gruppendaten ist das auf der Skala von 1 bis 7 trotz statistischer Signifikanz nur moderat angestiegene (!) Motivationslevel in Experiment 3 unproblematisch, da in Experiment 3 genauso wie in Experiment 1 keine systematischen Bedingungsunterschiede bei der gegebenen Mühe in Phase 2 auftraten.

## Lebenslauf

### **Dipl.-Psych. Sylvana Drewes**

Georg-August-Universität Göttingen  
Georg-Elias-Müller-Institut für Psychologie  
Abteilung Wirtschafts- und Sozialpsychologie  
Goßlerstraße 14  
37073 Göttingen



### **Persönliches**

Geburtsdatum: 26. Juni, 1982  
Geburtsort: Bad Gandersheim  
Staatsangehörigkeit: Deutsch

### **Beruflicher Werdegang**

2002	Abitur, Roswitha Gymnasium Bad Gandersheim
2002-2003	Studium der Wirtschaftspsychologie an der Fachhochschule Nordostniedersachsen, Lüneburg
2003-2008	Studium der Psychologie an der Georg-August-Universität Göttingen
2005-2006	Studium der Betriebswirtschaftslehre (B.Sc.; ohne Abschluss) an der Georg-August-Universität Göttingen
Seit 2008	Promotion im Fach Psychologie an der Georg-August-Universität Göttingen
2009-2012	Graduiertenstipendium der Konrad-Adenauer-Stiftung