

**Universität Ulm**

Institut für Business Analytics

**Algorithmen-Aversion im Controlling**

Experimente zur Wirkung des Arbeitsumfeldes  
auf die Algorithmen-Aversion in Forecasts

Dissertation

zur Erlangung des Doktorgrades Dr. rer. pol.

der Fakultät für Mathematik und Wirtschaftswissenschaften der Universität Ulm

vorgelegt von

**Markus Jung**

aus Herrenberg

2022

Amtierender Dekan: Prof. Dr. Stefan Funken

1. Gutachter: Prof. Dr. Mischa Seiter

2. Gutachter: Prof. Dr. Paul Wentges

Tag der Promotion: 19. Dezember 2022

# Inhaltsverzeichnis

<b>Abbildungsverzeichnis</b> .....	<b>VI</b>
<b>Tabellenverzeichnis</b> .....	<b>IX</b>
<b>Abkürzungsverzeichnis</b> .....	<b>XI</b>
<b>Zusammenfassung</b> .....	<b>XII</b>
<b>1 Einführung</b> .....	<b>1</b>
1.1 Problemstellung und Forschungsfrage .....	1
1.2 Relevanz der Forschungsfrage und Beitrag zur Literatur .....	6
1.3 Gang der Untersuchung .....	10
<b>2 Begriffsbestimmungen und Stand der Forschung</b> .....	<b>13</b>
2.1 Grundlagen zur Algorithmen-Aversion.....	14
2.1.1 Definition Algorithmus.....	15
2.1.2 Die Algorithmen-Aversion als Forschungsstrang .....	16
2.1.3 Theoretische Betrachtung der Algorithmen-Aversion .....	19
2.1.4 Einordnung wesentlicher Elemente der Algorithmen-Aversion in die Literatur.....	23
2.1.5 Stand der Forschung zur Algorithmen-Aversion .....	39
2.1.6 Zusammenfassung der bestehenden Erkenntnisse zur Algorithmen- Aversion.....	53
2.2 Einordnung der Algorithmen-Aversion in das Controlling .....	55
2.2.1 Die Bedeutung von Business Analytics für Algorithmen im Controlling	57
2.2.2 Einflüsse auf das Entscheidungsverhalten von Controllern .....	60
<b>3 Grundlagen zur empirischen Erhebung</b> .....	<b>68</b>
3.1 Grundlagen zu Experimenten .....	69
3.1.1 Experimente und Replikation .....	69
3.1.2 Plattform und Teilnehmer der Experimente .....	71

## Algorithmen-Aversion im Controlling

3.2	Experimente dieser Untersuchung.....	75
3.2.1	Forecastingaufgabe .....	78
3.2.2	Incentivierung .....	80
3.2.3	Algorithmus.....	80
3.2.4	Abgefragte Variablen.....	81
3.2.5	Besonderheiten durch die Replikation .....	83
3.3	Methodik zur Datenauswertung .....	83
<b>4</b>	<b>Studie 1.....</b>	<b>91</b>
4.1	Motivation .....	92
4.2	Hypothesenentwicklung.....	93
4.2.1	Algorithmen-Aversion .....	93
4.2.2	Der Einfluss von Zeitdruck auf die Algorithmen-Aversion.....	94
4.2.3	Der Einfluss eines „do your best“-Ziels auf die Algorithmen-Aversion ..	98
4.2.4	Der Einfluss von Entscheidungsrechten für Eingabeinformationen auf die Algorithmen-Aversion .....	100
4.2.5	Zusammenfassung der Hypothesen.....	102
4.3	Experiment und Beschreibung der Datengrundlage.....	103
4.3.1	Experimentaldesign und Operationalisierung der Manipulation.....	103
4.3.2	Deskriptive Statistik .....	108
4.4	Hypothesentests und Beurteilung der empirischen Ergebnisse.....	115
4.4.1	Hypothesentests.....	115
4.4.2	Zusatzanalysen und Diskussion H1 .....	117
4.4.3	Algorithmen-Aversion H2 bis H4.....	120
4.4.4	Schlussfolgerungen .....	127
4.4.5	Limitationen .....	130
<b>5</b>	<b>Studie 2.....</b>	<b>134</b>
5.1	Motivation .....	135
5.2	Hypothesenentwicklung.....	139
5.2.1	Vertiefung der theoretischen Grundlagen zu RPI .....	140

## Algorithmen-Aversion im Controlling

5.2.2	Der Einfluss von positiver RPI auf die Algorithmen-Aversion.....	151
5.2.3	Der Einfluss von negativer RPI auf die Algorithmen-Aversion .....	152
5.2.4	Zusammenfassung der Hypothesen .....	154
5.3	Experiment und Beschreibung der Datengrundlage .....	155
5.3.1	Experimentaldesign und Operationalisierung der Manipulation .....	155
5.3.2	Deskriptive Statistik .....	159
5.4	Hypothesentests und Beurteilung der empirischen Ergebnisse.....	167
5.4.1	Hypothesentests .....	167
5.4.2	Interpretationen der Ergebnisse und Zusatzanalysen.....	168
5.4.3	Schlussfolgerungen .....	179
5.4.4	Limitationen .....	182
<b>6</b>	<b>Erkenntnisfortschritt .....</b>	<b>185</b>
6.1	Algorithm Aversion Bias.....	185
6.2	Limitationen und Implikationen .....	187
6.2.1	Studienübergreifende Limitationen und künftiger Forschungsbedarf..	187
6.2.2	Praxisimplikationen .....	193
6.3	Abschließende Bewertung der Forschungsfrage .....	200
<b>7</b>	<b>Literaturverzeichnis.....</b>	<b>XIII</b>
<b>8</b>	<b>Anhang .....</b>	<b>XLV</b>
	<b>Eigenständigkeitserklärung .....</b>	<b>LIV</b>

## Abbildungsverzeichnis

Abbildung 1: Relevanz von KI und Algorithmen für den Wirtschaftsstandort Deutschland.....	7
Abbildung 2: Wissenschaftliche Disziplinen im Feld der Algorithmen-Aversion .....	8
Abbildung 3: Publikationen zur Algorithmen-Aversion von 2015 bis 2022 .....	8
Abbildung 4: Relevante Variablen aus dem Arbeitsumfeld von Forecastern.....	11
Abbildung 5: Aufbau der vorliegenden Arbeit.....	12
Abbildung 6: Aufbau von Kapitel 2 .....	13
Abbildung 7: TAM 1 nach Davis (1985).....	17
Abbildung 8: TAM 2 nach Venkatesh und Davis (2000).....	18
Abbildung 9: Algorithmen-Aversion nach Dietvorst et al. (2015) .....	21
Abbildung 10: Mediation zur Erklärung der Algorithmen-Aversion .....	22
Abbildung 11: Elemente der Algorithmen-Aversion.....	22
Abbildung 12: Forecasting als Element der Algorithmen-Aversion .....	23
Abbildung 13: Leistungsfeedback als Element der Algorithmen-Aversion .....	29
Abbildung 14: Confidence als Element der Algorithmen-Aversion .....	32
Abbildung 15: Zielsetzung und Incentivierung als Element der Algorithmen-Aversion .....	36
Abbildung 16: Die zweifache Bedeutung der Algorithmen-Aversion im Controlling ..	57
Abbildung 17: Aufgaben eines Data Scientists und eines Controllers.....	61
Abbildung 18: Incentive-Effort-Performance-Beziehung .....	63
Abbildung 19: Zuordnung der relevanten Variablen aus dem Arbeitsumfeld zu den empirischen Studien dieser Untersuchung .....	67
Abbildung 20: Aufbau von Kapitel 3 .....	69
Abbildung 21: Filterfrage zur Teilnahme an ähnlichen Experimenten .....	76
Abbildung 22: Entscheidung im incentivierten Forecast.....	77
Abbildung 23: Aufgabenstellung im Experiment.....	78
Abbildung 24: Informationen zur Aufgabe des Experiments .....	79
Abbildung 25: Beispielhafte Darstellung der Information zu einem U.S.-Bundesstaat .....	79

## Algorithmen-Aversion im Controlling

Abbildung 26: Darstellung des Leistungsfeedbacks.....	80
Abbildung 27: Forecastinggenauigkeit des Algorithmus .....	81
Abbildung 28: Schematische Darstellung der Mediationsanalyse .....	88
Abbildung 29: Aufbau von Kapitel 4 .....	91
Abbildung 30: Potenzielle Entwicklung der Algorithmen-Aversion durch eine steigende Algorithmen-Appreciation.....	94
Abbildung 31: Schematische Darstellung des Ablaufs des Experiments .....	104
Abbildung 32: Wahrnehmung von unterschiedlich starkem Zeitdruck.....	106
Abbildung 33: Zeitdruck in der Algorithmen-Appreciation .....	107
Abbildung 34: Darstellung des „do your best“-Ziels .....	107
Abbildung 35: Möglichkeit der Anpassung der der Eingabeinformationen .....	108
Abbildung 36: Alter der Teilnehmer in Studie 1 .....	109
Abbildung 37: Bildungsniveau der Teilnehmer in Studie 1 .....	109
Abbildung 38: Algorithmen-Aversion im Vergleich zu Dietvorst et al. (2015) .....	117
Abbildung 39: Mediationsanalyse zur Confidence in den Algorithmus in H1.....	119
Abbildung 40: Reduktion der Algorithmen-Aversion H2–H4 gegenüber H1 .....	121
Abbildung 41: Algorithmen-Aversion innerhalb der Treatmentgruppen.....	121
Abbildung 42: Mediationsanalyse Zeitdruck.....	124
Abbildung 43: Aufbau von Kapitel 5 .....	135
Abbildung 44: Erster Hinweis zur RPI .....	155
Abbildung 45: Treatment für positive RPI.....	156
Abbildung 46: Treatment für negative RPI .....	156
Abbildung 47: Finale Darstellung der RPI .....	157
Abbildung 48: Zweistufige Erhebung der Experimentalgruppen in Studie 2.....	159
Abbildung 49: Alter der Teilnehmer in Studie 2 .....	160
Abbildung 50: Bildungsniveau der Teilnehmer in Studie 2 .....	160
Abbildung 51: Algorithmen-Aversion bei positiver RPI .....	169
Abbildung 52: Verteilung der Confidencewerte in der CA .....	170
Abbildung 53: Verteilung der Confidencewerte in der PA .....	170
Abbildung 54: Algorithmen-Aversion bei negativer RPI .....	172
Abbildung 55: Mediationsanalyse RPI (1/2) .....	177

## Algorithmen-Aversion im Controlling

Abbildung 56: Mediationsanalyse RPI (2/2) .....	177
Abbildung 57: Aufbau von Kapitel 6 .....	185
Abbildung 58: Verlauf der Algorithmen-Aversion über die Häufigkeit des Leistungsfeedbacks .....	191

## Tabellenverzeichnis

Tabelle 1: Confidencewerte nach Dietvorst et al. (2015) (M und SD) .....	35
Tabelle 2: Modus für die Bonuszahlung .....	80
Tabelle 3: Abgefragte Variablen im Experiment.....	82
Tabelle 4: Zusammenfassung der Hypothesen in Studie 1 .....	102
Tabelle 5: Nomenklatur der Experimentalgruppen in Studie 1 .....	103
Tabelle 6: Ausfälle durch Zeitdruck im Pretest.....	105
Tabelle 7: Mittelwerte des wahrgenommenen Zeitdrucks (M und SD).....	106
Tabelle 8: Vergleich der Stichprobe aus Studie 1 mit Dietvorst et al. (2015) .....	110
Tabelle 9: Gruppeneinteilung in Studie 1 .....	110
Tabelle 10: Geschlecht und Alter nach Experimentalgruppen in Studie 1.....	111
Tabelle 11: Bildungsabschluss nach Experimentalgruppen in Studie 1 .....	111
Tabelle 12: Forecastinggenauigkeit von Forecastern und Algorithmus in Studie 1 (M und SD).....	112
Tabelle 13: Bonuszahlungen in \$ in Studie 1 (M und SD).....	113
Tabelle 14: Confidencewerte in Studie 1 (M und SD) .....	114
Tabelle 15: Entscheidung im incentivierten Forecast in Studie 1 .....	115
Tabelle 16: Hypothesentests Studie 1 .....	116
Tabelle 17: Kreuztabelle H1.....	116
Tabelle 18: Kreuztabelle H2.....	116
Tabelle 19: Kreuztabelle H3.....	117
Tabelle 20: Kreuztabelle H4.....	117
Tabelle 21: Algorithmen-Aversion innerhalb der Treatmentgruppen in Studie 1 .....	122
Tabelle 22: Confidencewerte innerhalb der Algorithmen-Aversion-Gruppen in Studie 1 (M) .....	122
Tabelle 23: Änderungen nach Übungsforecasts an der Auswahl der Eingabeinformationen (M und SD).....	126
Tabelle 24: Zusammenfassung der Hypothesentests in Studie 1 .....	127
Tabelle 25: Zusammenfassung der Hypothesen in Studie 2.....	154
Tabelle 26: Nomenklatur der Experimentalgruppen in Studie 2 .....	158

## Algorithmen-Aversion im Controlling

Tabelle 27: Vergleich der Stichproben Studie 1 und Studie 2 sowie Dietvorst et al. (2015) .....	161
Tabelle 28: Gruppengrößen zur Untersuchung zu positivem RPI .....	161
Tabelle 29: Gruppengrößen und Performance der Gruppen zu negativem RPI .....	162
Tabelle 30: Geschlecht und Alter nach Experimentalgruppen in Studie 2.....	162
Tabelle 31: Bildungsabschluss nach Experimentalgruppen in Studie 2 .....	163
Tabelle 32: Forecastinggenauigkeit von Forecastern und Algorithmus in Studie 2 (M und SD).....	163
Tabelle 33: Bonuszahlungen in \$ in Studie 2 (M und SD).....	164
Tabelle 34: Confidencewerte in Studie 2 (M und SD) .....	165
Tabelle 35: Einschätzung zum Algorithmeinsatz der Peers .....	166
Tabelle 36: Entscheidung im incentivierten Forecast in Studie 2 .....	167
Tabelle 37: Hypothesentests Studie 2 .....	167
Tabelle 38: Kreuztabelle H5.....	168
Tabelle 39: Kreuztabelle H6.....	168
Tabelle 40: Confidencewerte zu H5 (M und SD).....	169
Tabelle 41: Performancevergleich CA und PA (M und SD).....	171
Tabelle 42: Vergleich der Confidencewerte bei CBP und NBP (M und SD).....	173
Tabelle 43: Vergleich der Forecastinggenauigkeit CBP und NBP (M und SD) .....	174
Tabelle 44: Vergleich der Confidencewerte NGP und NBP (M und SD) .....	174
Tabelle 45: Vergleich der Confidencewerte CGP und CBP (M und SD) .....	174
Tabelle 46: Chi-Quadrat-Test zur Algorithmen-Aversion bei guter und schlechter Performance .....	175
Tabelle 47: Zusammenfassung der Hypothesentests in Studie 2 .....	179
Tabelle 48: Limitationen und künftiger Forschungsbedarf.....	188
Tabelle 49: Übersicht aller Hypothesentests dieser Arbeit.....	201

## Abkürzungsverzeichnis

FIT	Feedback Intervention Theory
HC-Standardfehler	Heteroskedastizität-konsistente Standardfehler
ICV	Internationaler Controller Verein e. V.
JAS	Judge-Advisor-System
KI	Künstliche Intelligenz
M	Mittelwert
MTurk	Amazon Mechanical Turk
MTurker	Amazon Mechanical Turk Worker
RPI	Relative Performance Information
SCT	Social Comparison Theory
SD	Standardabweichung
TAM	Technology Acceptance Model
WOA	Weight-on-Advice

### Zusammenfassung

Forecasts sind ein wesentliches Element des Controllings um Managemententscheidungen zu unterstützen. Durch neue Datenquellen und neue Techniken zur Datenanalyse sollen dabei zunehmend Algorithmen in Forecasts eingesetzt werden. Obwohl Algorithmen in der Regel präziser sind als Forecaster, werden sie von diesen häufig nicht eingesetzt. Eine solche Algorithmen-Aversion tritt insbesondere dann auf, wenn Forecaster mit Algorithmen interagieren und damit ein Feedback zur Leistung des Algorithmus erhalten. Dieses Leistungsfeedback demonstriert Forecastern, dass Algorithmen eine höhere Forecastinggenauigkeit als sie selbst haben, jedoch auch vom Algorithmus keine perfekten Forecasts erstellt werden. Da Forecasts einen Blick in die Zukunft richten, sind sie von Unsicherheit geprägt und perfekte Forecasts sind auch für Algorithmen unerreichbar. Die Algorithmen-Aversion wird demnach zu einer omnipräsenten Herausforderung im Controlling erwachsen.

In der Literatur wird die Algorithmen-Aversion bereits als robustes Phänomen beschrieben, es fehlt bislang jedoch an einem erweiterten Verständnis dafür, wie ausgeprägt sie im täglichen Arbeitsumfeld von Forecastern ist. Um dies näher zu ergründen wird in der vorliegenden Untersuchung der Forschungsfrage gefolgt, wie sich das Arbeitsumfeld von Forecastern auf die Algorithmen-Aversion auswirkt. Dazu werden vier besonders relevante Variablen des Arbeitsumfelds aus der Controlling-Literatur abgeleitet. Diese werden in zwei Experimentalstudien mit insgesamt 3243 Teilnehmern untersucht. In Studie 1 werden die Variablen *Zeitdruck*, „do your best“-Ziele und *Entscheidungsrechte, für Eingabeinformationen die der Algorithmus verarbeitet*, betrachtet. In Studie 2 liegt der Fokus auf relativem Leistungsfeedback zu Kollegen, sogenannter *Relative Performance Information* (RPI).

Die empirischen Ergebnisse zeigen, dass die Algorithmen-Aversion auch in dieser Untersuchung robust ist. Sie variiert jedoch in ihrer Ausprägung in Abhängigkeit der Variablen des Arbeitsumfelds. Hier zeigt die vorliegende Untersuchung, dass Zeitdruck die Algorithmen-Aversion absenkt. Weiter, dass RPI einen bedingten Einfluss auf die Algorithmen-Aversion hat. Positive RPI erhöht die Algorithmen-Aversion. Negative RPI hat bei guter Leistung eines Forecasters keinen Einfluss, bei schlechter Leistung jedoch, senkt sie die Algorithmen-Aversion. Für „do your best“-Ziele und Entscheidungsrechte, für Eingabeinformationen die der Algorithmus verarbeitet, werden keine Auswirkungen auf die Algorithmen-Aversion festgestellt.

# 1 Einführung

## 1.1 Problemstellung und Forschungsfrage

Für das Controlling eröffnen sich durch die digitale Transformation neue Potenziale.<sup>1</sup> Einerseits verbessert sich die Entscheidungsgrundlage durch in großen Mengen verfügbare Daten.<sup>2</sup> Andererseits werden intelligente Instrumente wie Algorithmen zur Datenverarbeitung zunehmend verfügbar und gewinnen stetig an Leistungsfähigkeit.<sup>3</sup> Folglich geht die steigende Menge an Daten mit immer effektiveren Möglichkeiten der Datenverarbeitung einher.<sup>4</sup>

Besonders große Potenziale erwachsen dabei im Forecasting – denn die Datenverfügbarkeit und die effektive Datenverarbeitung determinieren die Qualität eines Forecasts.<sup>5</sup> Forecasts haben den Charakter, unter Unsicherheit einen Wert oder Zustand in der Zukunft vorherzusagen.<sup>6</sup> Zur Verbesserung der Genauigkeit und zum Abbau der Unsicherheit können Forecaster<sup>7</sup> Entscheidungen durch Algorithmen unterstützen.<sup>8</sup>

In der Erstellung von Forecasts sind Algorithmen präziser als Menschen.<sup>9</sup> Dies gilt selbst für einfache Formen von Algorithmen wie lineare Modelle.<sup>10</sup> Aufgrund der überlegenen Leistung von Algorithmen in Bezug auf die Forecastinggenauigkeit ist demnach zu erwarten, dass Forecaster Algorithmen in ihre täglichen Arbeitsabläufe integrieren.<sup>11</sup> Die Literatur zeigt hingegen eine Vielzahl von Fällen, in denen Forecas-

---

<sup>1</sup> Vgl. Matt et al. 2016, S. 103; Quattrone 2016, S. 118; Schläfke et al. 2013, S. 110.

<sup>2</sup> Vgl. Al-Htaybat und Alberti-Alhtaybat 2017, S. 850.

<sup>3</sup> Vgl. Appelbaum et al. 2017, S. 29; Griffin und Wright 2015, S. 377.

<sup>4</sup> Vgl. Rikhardsson und Yigitbasioglu 2018, S. 37.

<sup>5</sup> Vgl. Butler und Ghosh 2015, S. 33; Chen et al. 2015, S. 1514; Utley 2014, S. 131.

<sup>6</sup> Vgl. Appelbaum et al. 2017, S. 32; Fildes et al. 2009, S. 3.

<sup>7</sup> Forecastingaufgaben werden in unterschiedlichen Berufen (u. a. Controlling, Produktion, Vertrieb) oder Aufgabenfeldern (u. a. rollierende Planung, vorausschauende Instandhaltung, Absatzplanung) durchgeführt. Es wird deshalb im Folgenden, stellvertretend für Menschen, die die Forecasts erstellen, der Begriff des Forecasters eingeführt. Der Fokus dieser Arbeit liegt darüber hinaus auf Forecastern im Controlling.

<sup>8</sup> Vgl. Seiter 2019, S. 25.

<sup>9</sup> Vgl. Dietvorst und Bharti 2020, S. 1302; Fildes et al. 2009, S. 3. Die Überlegenheit von Algorithmen hat dabei über unterschiedliche Domänen bestanden (siehe beispielsweise Grove et al. 2000, S. 19 oder Kuncel et al. 2013, S. 1060). Darüber hinaus sind Algorithmen in der Lage, selbst einen Experten, mit dessen Wissen sie angelernt wurden, zu übertreffen (Camerer 1981, S. 411).

<sup>10</sup> Vgl. Dawes 1979, S. 579.

<sup>11</sup> Vgl. Dietvorst et al. 2015, S. 114; Logg et al. 2019, S. 91; Venkatesh und Davis 2000, S. 190.

## Einführung

ter überlegene Algorithmen auf Kosten der Forecastinggenauigkeit und damit der Entscheidungsqualität ablehnen.<sup>12</sup> Dieses Phänomen wird in der jüngeren Literatur als *Algorithmen-Aversion* beschrieben.<sup>13</sup>

Die prägende Studie zur Algorithmen-Aversion stammt von Dietvorst et al. (2015). Sie definieren die Algorithmen-Aversion durch: „People Erroneously Avoid Algorithms after Seeing Them Err“<sup>14</sup>. Hinter dieser kurzen Definition steckt ein komplexes und weitreichendes Phänomen. Die Ablehnung von Algorithmen beruht im Wesentlichen auf der Beobachtung der Ergebnisse eines Algorithmus – also auf einem Leistungsfeedback zum Algorithmus. Wenn hieraus eine Ablehnung eines Algorithmus entsteht, ist dies aus zweifacher Sicht bedenklich: Erstens entstehen Forecasts in der Regel, um zukünftige Ereignisse vorherzusagen, weshalb es keine perfekten Forecasts geben kann. Die Ursache der Algorithmen-Aversion ist demnach im Forecasting omnipräsent. Zweitens konstatieren Dietvorst et al. (2015): „The aversion to algorithms is costly, not only for the participants in our studies who lost money when they chose not to tie their bonuses to the algorithm, but for society at large“<sup>15</sup>. Es werden durch eine Algorithmen-Aversion suboptimale Entscheidungen getroffen, die zu weitreichenden Wettbewerbsnachteilen führen können. Da die Leistungsfähigkeit von Algorithmen stetig wächst, führt die Aversion zu immer größeren Wettbewerbsnachteilen auf individueller genauso wie auf unternehmensweiter und gesellschaftlicher Ebene.<sup>16</sup> Da die Ablehnung von Algorithmen als ein systematischer Fehler im Entscheidungsprozess von Forecastern gesehen werden kann, bedarf es einer genauen Analyse der Ursachen, um die negativen Auswirkungen in Zukunft reduzieren zu können.<sup>17</sup>

Erste empirische Ergebnisse zeigen, dass die Stärke der Algorithmen-Aversion unterschiedlichen Einflussfaktoren unterliegen kann. Beispielsweise zeigen Logg et al. (2019) eine hohe Akzeptanz von Forecastern gegenüber neu eingeführten Algorithmen, also solchen, für die kein Leistungsfeedback vorliegt. Dies bezeichnen sie als *Algorithm Appreciation*.<sup>18</sup> Weiter zeigen Dietvorst et al. (2018), dass die Algorithmen-

---

<sup>12</sup> Vgl. Castelo et al. 2019; Dietvorst et al. 2015; Grove et al. 2000; Prah und van Swol 2017.

<sup>13</sup> Vgl. Dietvorst et al. 2018, 2015; Logg et al. 2019; Prah und van Swol 2017.

<sup>14</sup> Dietvorst et al. 2015, S. 114.

<sup>15</sup> Dietvorst et al. 2015, S. 124.

<sup>16</sup> Vgl. Fildes et al. 2009, S. 4.

<sup>17</sup> Vgl. Logg et al. 2019, S. 92.

<sup>18</sup> Vgl. Logg et al. 2019, S. 90.

## Einführung

Aversion gesenkt wird, wenn Forecaster eine Anpassung am Forecast eines Algorithmus vornehmen können.<sup>19</sup> Ferner ist die Aversion gegenüber Algorithmen in Aufgaben, die als subjektiv wahrgenommen werden, größer als in Aufgaben, die als objektiv wahrgenommen werden.<sup>20</sup> Castelo et al. (2019) sprechen daher davon, dass die Aversion abgebaut werden kann, indem Aufgaben einen objektiveren Charakter bekommen.<sup>21</sup> Insgesamt wird die Algorithmen-Aversion als robustes Phänomen beschrieben, das jedoch verschiedenen Einflussfaktoren unterliegt. Hieraus erwächst ein zunehmend komplexes Feld, über das bislang noch wenig empirisches Wissen besteht<sup>22</sup>, weshalb Logg et al. (2019) resümieren, „[...] algorithm aversion is not as straightforward as prior literature suggests, nor as contemporary researchers predict“<sup>23</sup>.

Eine besonders stark von der Algorithmen-Aversion betroffene betriebswirtschaftliche Funktion ist das Controlling.<sup>24</sup> Das Controlling unterliegt einem Wandel, der wesentlich durch den signifikant ansteigenden Einsatz von Daten und der entsprechenden Datenauswertung getrieben ist.<sup>25</sup> In der jüngeren Vergangenheit haben sich mehrere Ansätze zur Nutzung neuer Formen von Daten sowie neue Techniken zu deren Analyse durchgesetzt.<sup>26</sup> Controller und darunter insbesondere Forecaster sind zunehmend gezwungen, mit komplexen Algorithmen in Forecasts zu interagieren.<sup>27</sup> Dies wiederum wird durch die Einführung von Business Analytics im Controlling ermöglicht und sichergestellt.<sup>28</sup> Der Einsatz von Business Analytics wird zur Kernkompetenz im Controlling, denn „[...] verfügbare Daten so zu analysieren, dass betriebswirtschaftliche Probleme evidenzbasiert gelöst werden können, ist für Unternehmen eine zentrale Quelle von dauerhaften Wettbewerbsvorteilen“<sup>29</sup>. Das Ziel von Business Analytics im Controlling ist vordergründig die Erhöhung des Unternehmenserfolgs durch fakten-

---

<sup>19</sup> Vgl. Dietvorst et al. 2018, S. 1155.

<sup>20</sup> Vgl. Castelo et al. 2019, S. 809. Weitere Einflussfaktoren auf die Algorithmen-Aversion werden in Abschnitt 2.1.5 diskutiert.

<sup>21</sup> Vgl. Castelo et al. 2019, S. 809.

<sup>22</sup> Vgl. Burton et al. 2019, S. 1–3.

<sup>23</sup> Logg et al. 2019, S. 92.

<sup>24</sup> Vgl. Schäffer und Weber 2016, S. 8.

<sup>25</sup> Vgl. Kappes und Leyk 2018, S. 4; Quattrone 2016, S. 118.

<sup>26</sup> Vgl. Teoh 2018, S. 110.

<sup>27</sup> Vgl. Bergmann et al. 2020, S. 29; Dietvorst et al. 2018, S. 1155.

<sup>28</sup> Vgl. Appelbaum et al. 2017, S. 29; Gluchowski 2016, S. 273.

<sup>29</sup> Seiter 2019, S. 3.

## Einführung

und datenbasierte analytische Modelle in der Entscheidungsfindung.<sup>30</sup> Aus Kompetenzen im Business Analytics resultieren eine verbesserte Situationswahrnehmung und -erkennung, die Unterstützung bei strategischen Entscheidungen sowie die Optimierung von Forecasts durch effiziente Datenverarbeitung.<sup>31</sup>

Neue Forecasting-Ansätze sind demnach eine Manifestation der digitalen Transformation. Dabei ist zu berücksichtigen, dass diese in ein bestehendes Arbeitsumfeld von teilweise sehr erfahrenen Mitarbeitern integriert werden müssen.<sup>32</sup> Das bestehende Arbeitsumfeld ist historisch gewachsen und definiert die Prozesse und Routinen von Controllern.<sup>33</sup> Diese spiegeln sich wiederum in deren Verhalten wider, weshalb die Forschung im Controlling an dieser Komplexität ansetzen muss. Luft (2016) beschreibt Controller als „[...] social beings with complex and somewhat changeable motivations, not as isolated operators of stable (probably profit-maximizing) decision models“<sup>34</sup>. Der Prozess der Entscheidungsfindung in Forecasts ist demnach eingebettet in das bestehende Arbeitsumfeld, das wiederum maßgeblich das Verhalten beeinflusst.<sup>35</sup> Dabei können mehrere Einflüsse identifiziert werden, die in der Forschung zur Algorithmen-Aversion bisher keine Berücksichtigung gefunden haben. Beispielsweise sind hier Fristen zur Fertigstellung eines Forecasts zu nennen. Daten können kurzfristig eintreffen und müssen schnell verarbeitet werden, was in der Regel zu Zeitdruck beim Forecaster führt.<sup>36</sup> Es werden in Untersuchungen zur Algorithmen-Aversion häufig Incentives eingesetzt, die sich an der Qualität eines Forecasts orientieren.<sup>37</sup> Die Vergütung ist in der Praxis zumeist nicht an die Genauigkeit einzelner Forecasts gekoppelt, weshalb die Arbeitsanweisung in Unternehmen eher einer bestmöglichen Zielerreichung ähnelt („do your best“-Ziel).<sup>38</sup> Außerdem werden Algorithmen häufig in bestehenden Aufgaben bzw. Forecasts unterstützend eingeführt. Folglich ist Forecastern die Datengrundlage bekannt und sie können entscheiden, mit welchen Daten der Algorithmus arbeitet.<sup>39</sup> Zuletzt sind die Kollegen zu nennen. Im Controlling arbeitet eine

---

<sup>30</sup> Vgl. Mehanna et al. 2018, S. 502.

<sup>31</sup> Vgl. Buschbacher und Weber 2017, S. 54.

<sup>32</sup> Vgl. Matanovic et al. 2019, S. 54.

<sup>33</sup> Vgl. Günther et al. 2021, S. 45–55.

<sup>34</sup> Luft 2016, S. 9.

<sup>35</sup> Die folgende Aufzählung von Einflussfaktoren orientiert sich an Bonner 1999; Bonner und Sprinkle 2002. Die Auswahl und detaillierte Beschreibung dieser Einflüsse findet sich in Abschnitt 2.2.

<sup>36</sup> Vgl. Camerer et al. 2004, S. 862; Spiliopoulos und Ortmann 2018, S. 383.

<sup>37</sup> Vgl. Dietvorst et al. 2018, S. 1161, 2015, S. 119.

<sup>38</sup> Vgl. Webb et al. 2013, S. 1433.

<sup>39</sup> Vgl. Fildes et al. 2009, S. 3.

## Einführung

Vielzahl an Forecastern und die Genauigkeit ihrer Forecasts – also die Leistung – steht auch in Relation zur Leistung ihrer Kollegen.<sup>40</sup> Die Integration eines solchen Leistungsfeedbacks in Relation zu Kollegen wird als Peer-Feedback bzw. Relative Performance Information (RPI) beschrieben.

Vor dem Hintergrund dieser Einflüsse sollte sich der Erkenntnisfortschritt mit der Frage beschäftigen, wie die Algorithmen-Aversion im Arbeitsumfeld von Forecastern beeinflusst wird.<sup>41</sup> Unter der Berücksichtigung, dass es Umstände gibt, die die Algorithmen-Aversion reduzieren<sup>42</sup>, sprechen Möller et al. (2020) davon: „Which behavioral biases can be mitigated or can arise with the use of digital technologies?“<sup>43</sup> Da Forecasts eine der Kernaufgaben im Controlling sind und damit seit jeher eine etablierte Aufgabe darstellen, muss die Fragestellung zur Algorithmen-Aversion im Controlling grundlegend mit Bezug auf die Auswirkungen des Arbeitsumfelds erfolgen. Unterschiedliche Aufgabencharakteristika oder Steuerungsmechanismen zur Senkung der Algorithmen-Aversion werden stets mit dem bestehenden Arbeitsumfeld interagieren. Bevor diese weiterentwickelt werden können, muss die Einführung von Algorithmen in bestehende Aufgaben bei gegebenen personellen Ressourcen – also in das bestehende Arbeitsumfeld – verstanden werden. Es wird dazu in dieser Arbeit die übergeordnete Forschungsfrage gestellt:

### **Wie wirkt sich das Arbeitsumfeld von Forecastern auf die Algorithmen-Aversion aus?**

Um das Arbeitsumfeld von Forecastern hinsichtlich ihrer Algorithmen-Aversion systematisch zu untersuchen, wird auf der Arbeit von Bonner und Sprinkle (2002) aufgebaut. Bonner und Sprinkle (2002) spannen ein Framework auf, das Variablen aus dem Arbeitsumfeld von Controllern identifiziert, die von hoher Relevanz für die Erstellung von Forecasts sind. Diese Arbeit ist weitreichend in der Controlling-Literatur etabliert und findet breite Anwendung in der Forschung.<sup>44</sup>

---

<sup>40</sup> Vgl. Hannan et al. 2013, S. 553; Holderness et al. 2019, S. 1201.

<sup>41</sup> Vgl. Möller et al. 2020, S. 3.

<sup>42</sup> Siehe hierzu vor allem Castelo et al. 2019, S. 809; Dietvorst et al. 2018, S. 1155.

<sup>43</sup> Möller et al. 2020, S. 3.

<sup>44</sup> Siehe dazu beispielsweise Christ et al. 2016, S. 29; Eyring et al. 2021, S. 667; Feichter und Grabner 2020, S. 155; Hannan et al. 2008, S. 894; Lourenco 2019, S. 18; Malmi und Brown 2008, S. 293; Merchant und Otley 2006, S. 791; Speckbacher und Wabnegg 2020, o. S.; Tafkov 2013, S. 328; Verbeeten 2008, S. 428.

## Einführung

Die aus dem Framework abgeleiteten und damit für die vorliegende Arbeit relevanten Variablen sind: *Zeitdruck*, „*do your best*“-Ziele, *Entscheidungsrechte* und *RPI*.<sup>45</sup> Aus der Integration dieser vier Variablen aus dem Arbeitsumfeld in die Interaktion mit Algorithmen in Forecasts werden die Unterfragen abgeleitet.

Wie verhält sich die Algorithmen-Aversion in Forecasts, wenn ein Forecaster:

- unter Zeitdruck steht?
- ein „do your best“-Ziel hinsichtlich der Forecastinggenauigkeit zugewiesen bekommt?
- Entscheidungsrechte über die Daten hat, die der Algorithmus verarbeitet?
- RPI erhält?

### 1.2 Relevanz der Forschungsfrage und Beitrag zur Literatur

Die Forschungsfrage und die notwendigen Unterfragen sind sowohl aus praktischer wie auch wissenschaftlicher Sicht von hoher Relevanz. Die praktische Relevanz der Forschungsfrage findet sich in der 2018 von der Bundesregierung ausgerufenen *Strategie Künstliche Intelligenz* wieder.<sup>46</sup> Die aus der Künstlichen Intelligenz (KI) und damit aus Algorithmen erwachsende wirtschaftliche und gesellschaftliche Bedeutung wird in drei Dimensionen gesehen (siehe Abbildung 1). Durch den konsequenten Einsatz von Algorithmen wird ein Wertschöpfungspotenzial von über 430 Mrd. € bis 2030 prognostiziert.<sup>47</sup> Der Einsatz im Bereich *Forecasting* wird dabei als besonders hoch angesehen.<sup>48</sup> Algorithmen, die in der Praxis zum prädiktiven Arbeiten<sup>49</sup> eingesetzt werden, beruhen auf Technologien wie Analyseplattformen, In-Memory-Computing und Tools für KI, einschließlich Machine Learning und Robotic Process Automation.<sup>50</sup> Die große Herausforderung in der Praxis und der Schlüssel zur Verbesserung von Forecasts,

---

<sup>45</sup> Vgl. Bonner und Sprinkle 2002, S. 325–331.

<sup>46</sup> Die Bundesregierung nutzt den Begriff der KI dabei, einerseits Synonym für Algorithmen andererseits werden Algorithmen als integraler Bestandteil von KI gesehen. Im Verständnis der vorliegenden Arbeit können Algorithmen einfache mathematische Modelle sein aber auch intelligente Modelle, die zur Selbstoptimierung fähig sind. Zweitens sind der elementare Baustein für KI. Damit ist im Verständnis dieser Arbeit KI eine Anwendungstechnologie, die diejenigen Algorithmen nutzt, die zur Selbstoptimierung fähig sind. Vgl. Die Bundesregierung 2018.

<sup>47</sup> Vgl. PwC Europe 2018, S. 4.

<sup>48</sup> Vgl. Deloitte 2018, S. 6; Horváth & Partners 2017, S. 2.

<sup>49</sup> Siehe hierzu Abschnitt 2.1.1.

<sup>50</sup> Vgl. Deloitte 2018, S. 3–4.

## Einführung

wird in der Interaktion von Forecastern und Algorithmen gesehen.<sup>51</sup> Repetitive Aufgaben und solche mit hohem Datenvolumen werden von Algorithmen effektiver verarbeitet als von Forecastern. Der Arbeitsinhalt von Forecastern verschiebt sich in Richtung der Auswahl geeigneter Algorithmen in den jeweiligen Forecast, weshalb der Akzeptanz von Algorithmen eine hohe Bedeutung zukommt.<sup>52</sup>



Abbildung 1: Relevanz von KI und Algorithmen für den Wirtschaftsstandort Deutschland<sup>53</sup>

Die Grundlagen zum Verständnis der Algorithmen-Aversion wurden von Dietvorst et al. (2015) gelegt. Diese damit noch recht junge Arbeit erfreut sich bereits einer weiten Verbreitung mit großem Einfluss auf Wissenschaft und Praxis. Die Erkenntnisse werden daher auch als *influential* beschrieben.<sup>54</sup> Sie finden Anklang sowohl in der Forschung<sup>55</sup> als auch in der Populärwissenschaft<sup>56</sup>. Darüber hinaus zeigt sich ein Interesse an einem differenzierten Verständnis und der Weiterentwicklung der Erkenntnisse in unterschiedlichen Fachbereichen, wie dem Management (beispielsweise Marketing<sup>57</sup> oder Finance<sup>58</sup>) der Psychologie<sup>59</sup> oder der Informatik<sup>60</sup>. Ein ausdifferenziertes Bild der unterschiedlichen wissenschaftlichen Disziplinen, die seit 2015 die Forschung im Bereich der Algorithmen-Aversion voranbringen, findet sich in Abbildung 2. Die bedeutsamsten Teilgebiete sind demnach entweder betriebswirtschaftlich orientiert oder stammen aus der Psychologie sowie der Informatik.

<sup>51</sup> Vgl. Möller et al. 2020, S. 3.

<sup>52</sup> Vgl. Jung et al. 2018, S. 2; McKinsey 2017, S. 9.

<sup>53</sup> Eigene Darstellung nach Die Bundesregierung 2018, S. 4.

<sup>54</sup> Vgl. Logg et al. 2019, S. 90.

<sup>55</sup> Vgl. unter anderem Castelo et al. 2019; Efendić et al. 2020; Logg et al. 2019.

<sup>56</sup> Vgl. unter anderem Frick 2015; Harrell 2016.

<sup>57</sup> Vgl. Castelo et al. 2019.

<sup>58</sup> Vgl. Filiz et al. 2021; Germann und Merkle 2019.

<sup>59</sup> Vgl. Dietvorst et al. 2015; Efendić et al. 2020.

<sup>60</sup> Vgl. Feng und Gao 2020; Renier et al. 2021.

## Einführung



Abbildung 2: Wissenschaftliche Disziplinen im Feld der Algorithmen-Aversion<sup>61</sup>

Das große Interesse in wissenschaftlichen Publikationen wird durch den raschen Anstieg an Publikationen zur Algorithmen-Aversion in Abbildung 3 deutlich. Dabei zeigt Abbildung 3 die jährlichen Veröffentlichungen zum Schlagwort *Algorithm Aversion*. Bis in das Jahr 2020 kann von einer überschaubaren Anzahl an Publikationen berichtet werden. Diese sind daher prägend für den Stand der Forschung und wurden in den Folgejahren intensiv zitiert und aufgegriffen. Die relevantesten und wissenschaftlich fundierten Arbeiten sind: Dietvorst et al. (2015) (322 Zitationen), Prahl und van Swol (2017) (39), Dietvorst et al. (2018) (129), Burton et al. (2019) (49), Castelo et al. (2019) (92) und Logg et al. (2019) (135).<sup>62</sup>

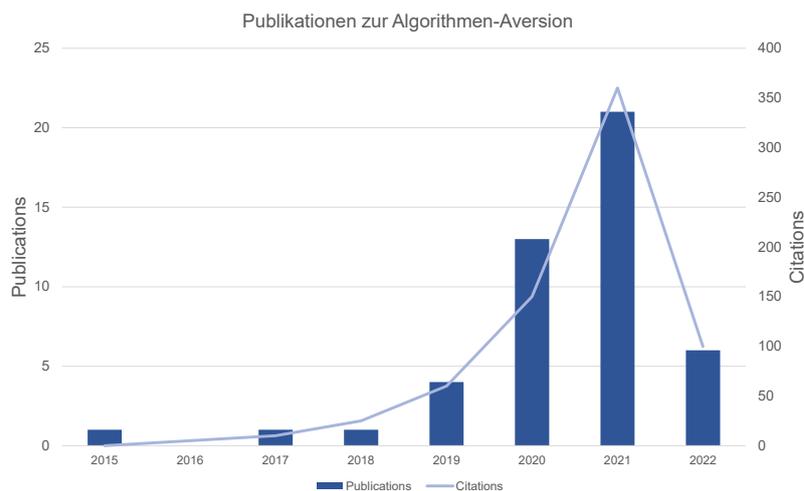


Abbildung 3: Publikationen zur Algorithmen-Aversion von 2015 bis 2022<sup>63</sup>

<sup>61</sup> Eigene Darstellung. Die Größe der Kacheln spiegelt die Anzahl der Publikationen, die Farben spiegeln die Zuordnung zu den Fachbereichen wider (Management (Dunkelblau), Psychologie (Blau) und Informatik (Hellblau)).

<sup>62</sup> Zitationen basierend auf <https://www.webofscience.com/> (letzte Aktualisierung am 29.04.2022).

<sup>63</sup> Eigene Darstellung adaptiert aus <https://www.webofscience.com/> (letzte Aktualisierung am 29.04.2022).

## Einführung

Mit dem starken Anstieg der Publikationen geht eine Erweiterung des Verständnisses der Algorithmen-Aversion einher. Diese Erweiterung kann mangels Sorgfalt in der Interpretation und des Erkenntnisfortschritts als Verwässerung interpretiert werden.<sup>64</sup> Die Algorithmen-Aversion wird in einigen Quellen nicht mehr identisch dazu verwendet bzw. interpretiert, wie ursprünglich von Dietvorst et al. (2015) definiert und empirisch gezeigt. Dabei wird Algorithmen-Aversion häufig systematisch überschätzt und fast schon dramatisiert. Diese Überschätzung trägt zu einer künstlich erhöhten Relevanz der Forschung bei, weshalb es einer eindeutigen Abgrenzung bedarf. Prahl und van Swol (2017) interpretieren die Algorithmen-Aversion beispielsweise als „[...] humans are trusted more than computers“<sup>65</sup>, und führen keinen Verweis auf das unerlässliche Leistungsfeedback als Auslöser der Algorithmen-Aversion an.<sup>66</sup> Der starke Anstieg an Publikationen und die Freiheiten in der Ergebnisinterpretation zeigen, dass es nicht nur einer eindeutigen Definition sondern auch eines nachhaltigen Verständnisses der Algorithmen-Aversion bedarf. Hierzu konstatieren Döring und Bortz (2016) im Sinne der Falsifizierbarkeit von wissenschaftlichen Arbeiten: „[D]ie in der Theorie verwendeten theoretischen Konstrukte, die sich auf Merkmale der sozialen Wirklichkeit beziehen, müssen so eindeutig definiert sein, dass sie prinzipiell empirisch erfassbar bzw. messbar sind“<sup>67</sup>. Darüber hinaus gilt, dass für die Algorithmen-Aversion noch kein hoher empirischer Bewährungsgrad erfolgt ist, da eine rasche und zu breite Interpretation der Ergebnisse stattfindet.<sup>68</sup>

Um der (zu) breiten Verwendung der Algorithmen-Aversion in der aktuellen Forschung zu begegnen, werden in dieser Arbeit in Abschnitt 2.1 die Algorithmen-Aversion im engeren Sinne<sup>69</sup> und die Algorithmen-Aversion im weiteren Sinne – wortgetreu als die Ablehnung von Algorithmen – unterschieden.<sup>70</sup> Auf Basis dieses Verständnisses wird die Algorithmen-Aversion im Kontext des Arbeitsumfelds von Controllern be-

---

<sup>64</sup> Dies kann auch im Sinne eines *Publication Bias* interpretiert werden, da Ergebnisse publiziert werden die zwar Signifikanz besitzen, jedoch vom theoretischen Konstrukt abweichen. Siehe dazu Döring und Bortz 2016, S. 57.

<sup>65</sup> Prahl und van Swol 2017, S. 693.

<sup>66</sup> Zur Diskussion um die Fehlinterpretation der Algorithmen-Aversion siehe auch Logg et al. 2019, S. 91.

<sup>67</sup> Döring und Bortz 2016, S. 56.

<sup>68</sup> Vgl. Döring und Bortz 2016, S. 57.

<sup>69</sup> Maßgeblich hierfür ist Dietvorst et al. 2015, unterstützt durch Berger et al. 2021; Dietvorst et al. 2018; Efendić et al. 2020; Filiz et al. 2021; Logg et al. 2019.

<sup>70</sup> Diese Differenzierung folgt der notwendigen Eingrenzung des Untersuchungsgegenstandes und eindeutigen Begriffsbestimmung. Im Folgenden wird zur Vereinfachung zumeist von der Algorithmen-Aversion gesprochen, wobei darunter die Algorithmen-Aversion i. e. S. verstanden wird.

trachtet. Es werden mit der Algorithmen-Aversion und dem Controlling zwei Literaturstränge bzw. Forschungsfelder miteinander verflochten. Durch diese Dualität der Forschungsstränge wird ein Beitrag zum wissenschaftlichen Erkenntnisfortschritt in beiden Feldern geleistet.

Erstens werden Erkenntnisse zur Literatur der Algorithmen-Aversion gewonnen. Es wird zunächst gezeigt, wie robust die Algorithmen-Aversion auch in den 2020er Jahren ist (Abschnitt 4.4.2). Darüber hinaus zeigen die empirischen Ergebnisse, dass die Algorithmen-Aversion über die Variablen aus dem Arbeitsumfeld von Forecastern Bestand hat. Zweitens wird ein integraler Beitrag zur Literatur der Algorithmen-Aversion und der Controlling-Literatur geleistet. Die vier Variablen aus der Controlling-Literatur, die das Arbeitsumfeld von Forecastern näher spezifizieren, interagieren mit der Stärke der Algorithmen-Aversion: Hier zeigt die vorliegende Arbeit, dass Zeitdruck die Algorithmen-Aversion absenkt (Abschnitt 4.4.1). Weiter, dass RPI einen bedingten Einfluss auf die Algorithmen-Aversion hat. Positive RPI erhöht die Algorithmen-Aversion. Negative RPI hat bei guter Leistung eines Forecasters keinen Einfluss, bei schlechter Leistung jedoch senkt sie die Algorithmen-Aversion (Abschnitt 5.4.1). Für „do your best“-Ziele und Entscheidungsrechte für Eingabeinformationen des Algorithmus hingegen konnten keine Auswirkungen auf die Algorithmen-Aversion gezeigt werden (Abschnitt 4.4.1).

### **1.3 Gang der Untersuchung**

Aufbauend auf der Problemstellung und der Relevanz der Forschungsfrage in Kapitel 1 folgt Kapitel 2 der Dualität der Forschungsfrage. Um eine Antwort darauf geben zu können, wie sich das bestehende Arbeitsumfeld von Forecastern auf die Einführung von Algorithmen auswirkt, wird zunächst in Abschnitt 2.1 die Algorithmen-Aversion abgegrenzt und der Stand der Forschung aufgezeigt. Anschließend zielt Abschnitt 2.2 darauf ab, die Forschungslücke zur Algorithmen-Aversion im Controlling herauszuarbeiten. Dazu wird in Abschnitt 2.2 spezifiziert, wie Algorithmen Einzug in das Controlling nehmen und weshalb die vier zu untersuchenden Variablen aus dem Arbeitsumfeld von Forecastern als im Zusammenhang mit der Algorithmen-Aversion relevant klassifiziert werden müssen. Die vier Variablen sind in Abbildung 4 nochmals zusammengefasst.

## Einführung

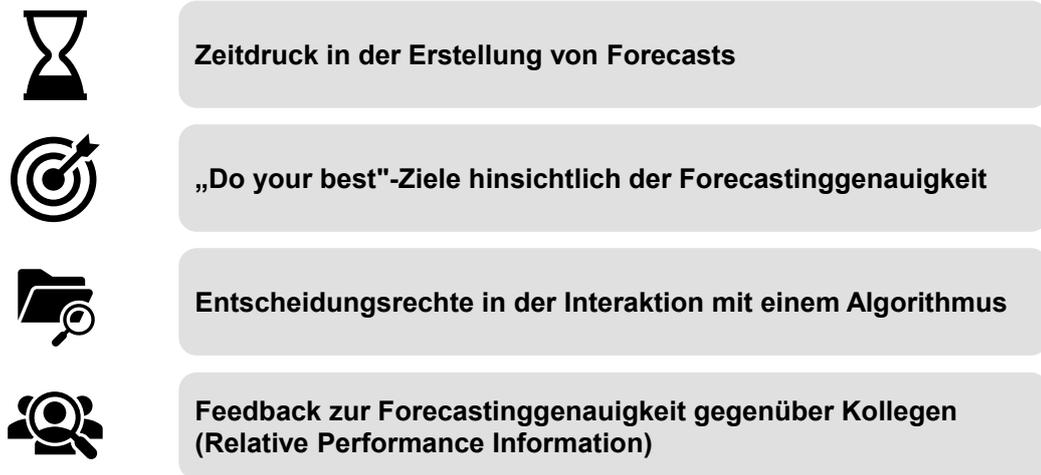


Abbildung 4: Relevante Variablen aus dem Arbeitsumfeld von Forecastern<sup>71</sup>

Um eine hohe Anschlussfähigkeit dieser vier Variablen aus dem Arbeitsumfeld von Forecastern zu gewährleisten, werden sie in ein etabliertes Experimentalsetting aus der Forschung zur Algorithmen-Aversion integriert.<sup>72</sup> In Kapitel 3 werden der Aufbau dieses Experiments und wesentliche Merkmale, wie die Incentivierung, die Stichprobe oder der zugrundeliegende Algorithmus, beschrieben. Kapitel 4 stellt Studie 1 dar, die den Einfluss von Zeitdruck, „do your best“-Zielen und Entscheidungsrechten auf die Algorithmen-Aversion untersucht. In Kapitel 5 wird Studie 2 vorgestellt, bei der die theoretischen Überlegungen zur RPI vertieft werden und ebenfalls untersucht wird, wie sich diese auf die Algorithmen-Aversion auswirkt. Abschließend wird in Kapitel 6 der Erkenntnisfortschritt zusammengefasst. Dies beinhaltet die Notwendigkeit der Definition eines Algorithm Aversion Bias. Ferner werden in Kapitel 6 die Limitationen aufgezeigt, denen diese Arbeit unterliegt. Zuletzt wird daraus der künftige Forschungsbedarf aufgedeckt und es werden Implikationen für die Praxis abgeleitet. Eine Übersicht zum Aufbau der vorliegenden Arbeit ist in Abbildung 5 dargestellt.

<sup>71</sup> Eigene Darstellung.

<sup>72</sup> Vgl. Dietvorst et al. 2015 Studie 3b.

# Einführung

Kapitel 1: Einführung	
Problemstellung und Relevanz der Forschungsfrage	Gang der Untersuchung
Kapitel 2: Begriffsbestimmungen und Stand der Forschung	
Grundlagen und Stand der Forschung zur Algorithmen-Aversion	Die Relevanz der Algorithmen-Aversion im Controlling
Kapitel 3: Grundlagen zur empirischen Erhebung	
Ablauf und Inhalt der Experimente	Methodische Grundlagen in der Datenauswertung
Kapitel 4: Studie 1	
Die Wirkung von - Zeitdruck - „do your best“-Zielen - Entscheidungsrechten auf die Algorithmen-Aversion	Hypothesenentwicklung
	Empirische Erhebung
	Diskussion und Limitationen
Kapitel 5: Studie 2	
Die Wirkung von - positiver RPI - negativer RPI auf die Algorithmen-Aversion	Hypothesenentwicklung
	Empirische Erhebung
	Diskussion und Limitationen
Kapitel 6: Erkenntnisfortschritt	
Der Algorithm Aversion Bias	Abschließende Bewertung mit Praxisimplikationen

Abbildung 5: Aufbau der vorliegenden Arbeit<sup>73</sup>

<sup>73</sup> Eigene Darstellung.

## 2 Begriffsbestimmungen und Stand der Forschung

Im Folgenden werden die Grundlagen für das Verständnis der Algorithmen-Aversion in dieser Arbeit herausgearbeitet. Abbildung 6 fasst den Aufbau von Kapitel 2 zusammen.

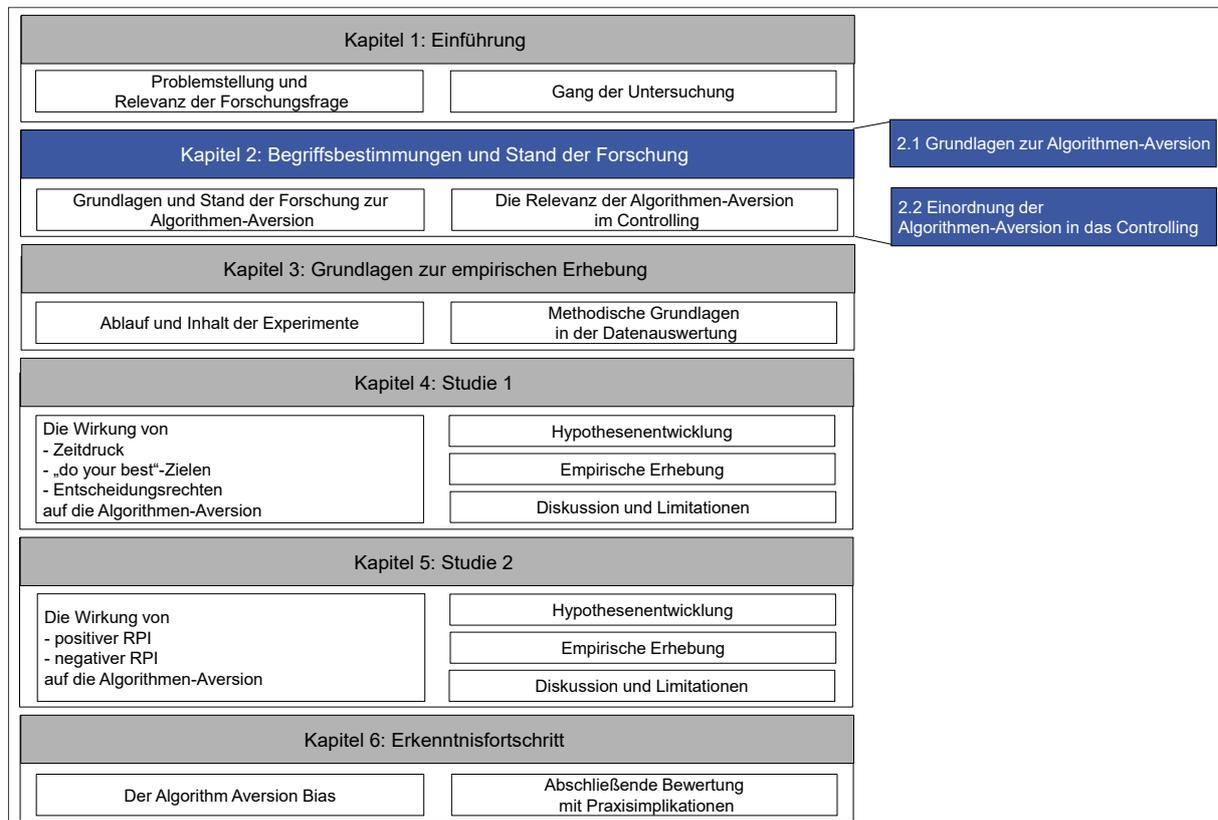


Abbildung 6: Aufbau von Kapitel 2<sup>74</sup>

In Abschnitt 2.1.1 wird der Begriff *Algorithmen-Aversion* definiert und dargelegt, was in dieser Arbeit unter einem *Algorithmus* verstanden wird. Außerdem wird in Abschnitt 2.1.2 die Algorithmen-Aversion von bestehenden Konzepten zur Erklärung der Interaktion von Forecastern mit Technologie abgegrenzt. Anschließend werden die wesentlichen Elemente der Algorithmen-Aversion detailliert beschrieben und Anknüpfungspunkte aus verwandten Forschungsgebieten dazu betrachtet (siehe Abschnitt 2.1.3 und Abschnitt 2.1.4). Die Forschung zur Algorithmen-Aversion selbst wurde aufbauend auf Dietvorst et al. (2015) weiterentwickelt. Diese Weiterentwicklung bedarf einer klaren Abgrenzung der sich innerhalb der Algorithmen-Aversion bildenden Erklärungsansätze (siehe Abschnitt 2.1.5). Anschließend wird die Algorithmen-Aversion in

<sup>74</sup> Eigene Darstellung.

Abschnitt 2.2 in das Controlling eingeordnet. Durch den tiefgreifenden Wandel im Controlling aufgrund der digitalen Transformation sind die Relevanz und damit auch die Akzeptanz von Algorithmen unbestreitbar. Da die Forschung zur Algorithmen-Aversion jedoch noch keine explizite Berücksichtigung im Controlling findet, wird sie über das *Business Analytics* eingeordnet (siehe Abschnitt 2.2.1). Hier kristallisiert sich das Arbeitsumfeld von Controllern als wesentlicher Betrachtungsrahmen heraus. Die besonders relevanten Aspekte aus diesem Arbeitsfeld werden daher in Abschnitt 2.2.2 literaturbasiert herausgearbeitet.

### 2.1 Grundlagen zur Algorithmen-Aversion

Die Algorithmen-Aversion steht nicht im Einklang mit einer rationalen Entscheidungsfindung wie sie die Agency-Theorie postuliert. Die Agency-Theorie trifft die Annahme, dass Individuen ihren Nutzen maximieren.<sup>75</sup> Eine Agency-Beziehung liegt dann vor, wenn eine oder mehrere Personen (Prinzipale) andere (Agenten) anstellen, um ihnen Aufgaben zu übertragen. Der Prinzipal überträgt dabei Rechte und Pflichten zur Erfüllung der Aufgabe an den Agenten. Es wird angenommen, dass Individuen rational handeln, ihre Nutzenfunktion kennen und diese maximieren möchten. Agenten optimieren ihre Handlungen innerhalb der ihnen übertragenen Rechte und Pflichten, um den höchsten Nutzen zu erreichen.<sup>76</sup> Der Nutzen des Agenten besteht meist aus zwei Größen: *Reichtum* und *Freizeit*, wobei Freizeit als *reduction in effort* verstanden werden kann.<sup>77</sup> Incentives sind dabei als eine Art extrinsische Motivation des Prinzipals an den Agenten zu sehen, um diesen zu einer bestimmten Leistung zu reizen, also seine Freizeit gegen Reichtum einzutauschen.<sup>78</sup> Zusammengefasst bedeutet dies: „[A]gency theory suggests that incentives play a fundamental role in motivation and the control of performance because individuals have utility for increases in wealth“<sup>79</sup>.

Unter diesen Prämissen steckt in der Algorithmen-Aversion ein zweifacher irrationaler Widerspruch. Zum einen könnte der *Reichtum* erhöht werden, wenn auf den Algorithmus zurückgegriffen wird, denn er hat die deutlich bessere Forecastinggenauigkeit und erreicht damit das höhere Incentive. Zum anderen kann die Anstrengung sowohl hinsichtlich der Intensität als auch hinsichtlich der Dauer reduziert werden,

---

<sup>75</sup> Vgl. Baiman 1990; Eisenhardt 1989.

<sup>76</sup> Vgl. Verbeeten 2008, S. 433.

<sup>77</sup> Vgl. Bonner und Sprinkle 2002, S. 308.

<sup>78</sup> Vgl. Verbeeten 2008, S. 433.

<sup>79</sup> Verbeeten 2008, S. 433.

wenn zügig eine Entscheidung zugunsten des Algorithmus gefällt wird. Im Folgenden wird daher näher beschrieben, wie die Algorithmen-Aversion erklärt wird und welche Faktoren sie beeinflussen.

### 2.1.1 Definition Algorithmus

Ein Algorithmus ist im Wesentlichen eine Abfolge von mathematischen Berechnungen mit einem bestimmten Ziel. In der Literatur zu *Computer Science* wird er beispielsweise nach Rogers (1987) definiert als „[...] an algorithm is a procedure for computing a function [...]“<sup>80</sup>. Analog dazu definiert Knuth (1990) einen Algorithmus als „[...] a finite set of rules which gives a sequence of operations for solving a specific type of problem [...]“<sup>81</sup>. Gemein ist den Definitionen ein regelbasiertes Vorgehen zur Lösung eines Problems bzw. die Erreichung eines bestimmten Ziels. Durch die Leistungsfähigkeit von modernen Computern können das regelbasierte Vorgehen und die Art der Probleme bzw. Ziele hinsichtlich Quantität und Komplexität weitreichend ausgelegt werden: Algorithmen können große, unstrukturierte Datensätze erfassen und auswerten.<sup>82</sup>

In der Forschung zur Algorithmen-Aversion wird der eingesetzte Algorithmus oberflächlich beschrieben. Es wird dadurch ein latentes, einheitliches Verständnis für einen Algorithmus vorausgesetzt und beim Forecaster wenig Fokus auf den Algorithmus selbst erzeugt. Dietvorst et al. (2015) und Dietvorst et al. (2018) sprechen bei einem Algorithmus in ihren Experimenten von „a sophisticated model, put together by thoughtful analysts“<sup>83</sup> und geben ferner den Hinweis, dass der Algorithmus die gleichen Daten nutzt, die auch den Forecastern zur Verfügung gestellt werden – und darüber hinaus keine Informationen hat. Es zeigt sich, dass die wesentlichen Informationen zum Algorithmus demnach die Daten, mit denen der Algorithmus arbeitet, die Herkunft und teilweise die Güte sind. Dabei ist zu erwähnen, dass die Güte explizit durch die Angabe einer Genauigkeit oder implizit durch Erfahrungswerte in der Zusammenarbeit mit einem Algorithmus entsteht.

Diese generischen Betrachtungen eines Algorithmus wurden von Logg et al. (2019) im Kontext der Algorithmen-Aversion evaluiert.<sup>84</sup> Vor Durchführung ihrer eigentlichen Untersuchung zur Algorithmen-Aversion bzw. Algorithmen-Appreciation, haben

---

<sup>80</sup> Rogers 1987, S. 2.

<sup>81</sup> Knuth 1990, S. 4.

<sup>82</sup> Vgl. Logg et al. 2019, S. 90; Vitale et al. 2020, S. 107.

<sup>83</sup> Dietvorst et al. 2018, S. 1157, 2015, S. 118.

<sup>84</sup> Vgl. Logg et al. 2019, S. 93.

sie die Teilnehmer eines Online-Experiments gebeten, in einem Freifeld ihre persönliche Definition eines Algorithmus zu beschreiben.<sup>85</sup> Knappe 70 % der Definitionen konnten mittels Kodierung zusammengefasst werden zu: „An algorithm is a set of equations to find an answer. It will spit out an answer“<sup>86</sup> bzw. „An algorithm is a systematic way of solving problems. It looks at a problem and goes through a process to figure out the solution“<sup>87</sup>. Dieses Verständnis von Algorithmen deckt sich weitestgehend mit der technischen Definition, wie sie in der vorliegenden Arbeit gegeben ist. Es ist daher davon auszugehen, dass die Probanden der Experimentalstudien der vorliegenden Arbeit ein einheitliches und der Computer Science deckungsgleiches Verständnis aufweisen.<sup>88</sup>

So kann festgehalten werden, dass in der Forschung zur Algorithmen-Aversion der Algorithmus und sein technischer Aufbau im Hintergrund stehen. Da bereits einfache lineare Modelle häufig genauere Forecasts als Forecaster treffen, wird es als nicht notwendig angesehen, zusätzliche Komplexität durch viele oder womöglich zu viele Informationen über einen Algorithmus zu erzeugen.

### 2.1.2 Die Algorithmen-Aversion als Forschungsstrang

Die Forschung zur Interaktion von Forecastern und Technologie ist weitläufig. Dies ergibt sich aus den unzähligen Arten von Automatisierungstechnologien. Neben physischen Robotern kann der Begriff in Hinblick auf beispielsweise Softwarelösungen im betrieblichen Umfeld oder beim autonomen Fahren nahezu beliebig ausgedehnt werden, denn je Technologie ergeben sich wiederum unterschiedliche Anwendungsszenarien<sup>89</sup> und damit unterschiedliche Bedingungen für den jeweiligen Anwender.<sup>90</sup> Dies wiederum resultiert in mehr oder weniger spezifischen Forschungsfeldern, die sich mit der Interaktion von Forecaster und Technologie beschäftigen.<sup>91</sup>

Ein dominierendes Konzept zur Untersuchung und Erklärung von Mensch-und-Technologieinteraktion ist das Technology Acceptance Model (TAM).<sup>92</sup> Im TAM wird

---

<sup>85</sup> Siehe dazu auch Abschnitt 2.1.5.4.

<sup>86</sup> Logg et al. 2019, S. 94.

<sup>87</sup> Logg et al. 2019, S. 94.

<sup>88</sup> Zu den Teilnehmer in den Studien der vorliegenden Arbeit siehe Abschnitt 3.1.2.

<sup>89</sup> Zum Beispiel in Navigationssystemen oder in Flugmanagement-Systemen. Sie werden mitunter zur Erfassung und Analyse von Informationen sowie zur Entscheidungsfindung verwendet. Siehe Hoff und Bashir 2015, S. 407.

<sup>90</sup> Vgl. Marangunić und Granić 2015, S. 81.

<sup>91</sup> Vgl. Lee und See (2004), S. 65.

<sup>92</sup> Vgl. Marangunić und Granić 2015, S. 81.

## Begriffsbestimmungen und Stand der Forschung

*Automation* weitläufig definiert mit: „Automation is technology that actively selects data, transforms information, makes decisions, or controls processes“<sup>93</sup>. Entsprechend erfassen Lee und See (2004) den Einsatzbereich einer solchen Technologie umfangreich, mit: „Sophisticated automation is becoming ubiquitous, appearing in work environments as diverse as aviation, maritime operations, process control, motor vehicle operation, and information retrieval“<sup>94</sup>. Diese Definitionen der Automation und ihrer Anwendungsfelder zeigen, wie umfassend das Forschungsfeld des TAM eingesetzt werden kann.

Das TAM hat daher im Laufe der Jahre einige Weiterentwicklungen erfahren. Es ist im Wesentlichen geprägt durch Davis (1985), Venkatesh und Bala (2008), Venkatesh und Davis (2000) und Venkatesh et al. (2003). Wesentliche Entwicklungen im TAM können differenziert werden durch das TAM 1 und das TAM 2. Dabei ist das TAM 1 sehr breit ausgerichtet (siehe Abbildung 7) und wurde aufgrund seiner einfachen Anwendung populär.<sup>95</sup> Es fokussiert die mediiierende Rolle von *Perceived Usefulness* und *Perceived Ease of Use* auf die Motivation eines Anwenders, eine bestimmte Automatisierungstechnologie einzusetzen.<sup>96</sup> Das relativ simple, dafür aber robuste Modell des TAM 1 ist zugleich dessen Schwachstelle, denn es werden eine Vielzahl potenzieller Einflussfaktoren, wie beispielsweise Jobrelevanz oder Erfahrung, nicht berücksichtigt.<sup>97</sup>

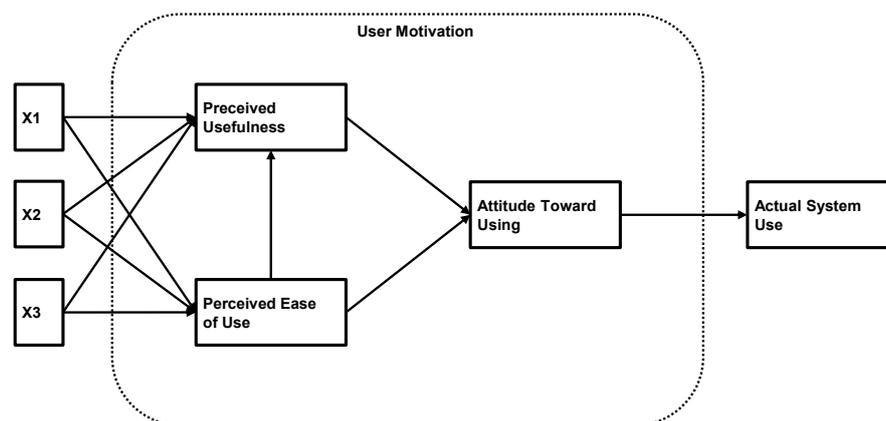


Abbildung 7: TAM 1 nach Davis (1985)<sup>98</sup>

<sup>93</sup> Lee und See 2004, S. 50.

<sup>94</sup> Lee und See 2004, S. 50.

<sup>95</sup> Vgl. Marangunić und Granić 2015, S. 81.

<sup>96</sup> Vgl. Davis 1985.

<sup>97</sup> Vgl. Chuttur 2009, S. 2.

<sup>98</sup> Eigene Darstellung nach Davis 1985, S. 24.

## Begriffsbestimmungen und Stand der Forschung

Mit dem TAM 2 wurde darauf reagiert und es kamen mehrere Einflussfaktoren, wie die Relevanz für die Arbeit bzw. den Beruf oder die Outputrelevanz als Prädiktoren in das Modell hinzu (siehe Abbildung 8).<sup>99</sup> Durch die Hinzunahme mehrerer solcher Prädiktoren wurde das TAM 2 wiederum unübersichtlich und einschränkend. Denn durch die Vorgabe vielerlei Einflussfaktoren wird die Untersuchung neuer, entdeckender Zusammenhänge erschwert.

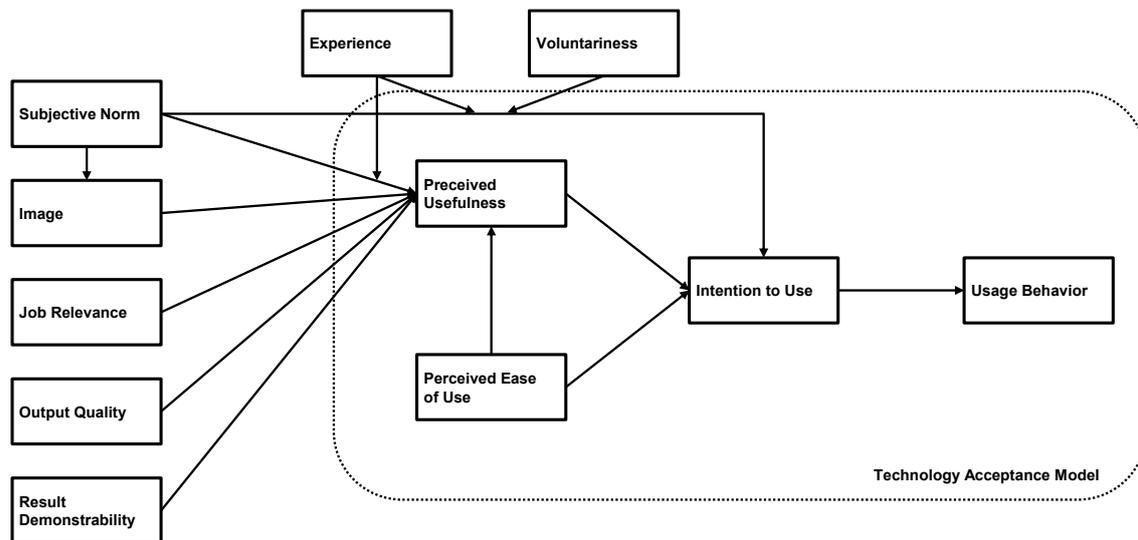


Abbildung 8: TAM 2 nach Venkatesh und Davis (2000)<sup>100</sup>

Zusammengefasst lässt sich aus der Entwicklung des TAM 1 und des TAM 2 erkennen, dass simple Modelle an ihre Grenzen stoßen, wenn sie einen spezifischen Kontext erklären sollen. Hingegen komplexe Modelle mit einem omnipotenten Erklärungsansatz keinen effektiven Erkenntnisfortschritt mehr zulassen. In einem solchen Spannungsfeld kann die Algorithmen-Aversion in Forecasts gesehen werden. Sie stellt einen speziellen Fall dar, in dem die Bedingungen für eine theoretische hohe Akzeptanz von Algorithmen gegeben sind, denn die Algorithmen liefern eine hohe Forecastinggenauigkeit, werden allerdings nicht eingesetzt. Daher wird die Algorithmen-Aversion als eigenständiger Ansatz zur Erklärung der Ablehnung von Algorithmen verwendet.

Im Folgenden werden, die Vorteile eines prägnanten Erklärungsmodells mit den technologiespezifischen Faktoren eines Algorithmus im Kontext von Forecasts kombiniert. Dadurch wird einerseits ein hoher Erklärungsgehalt ermöglicht und andererseits

<sup>99</sup> Vgl. Legris et al. 2003, S. 191.

<sup>100</sup> Eigene Darstellung nach Venkatesh und Davis 2000, S. 188.

werden bedeutsame Spezifika nicht außen vorgelassen. Mittels des Modells zur Algorithmen-Aversion wird der Untersuchungskontext auf die Entscheidungsfindung in Forecasts reduziert. Technologisch wird daher ein Forecast-Algorithmus fokussiert.<sup>101</sup>

### 2.1.3 Theoretische Betrachtung der Algorithmen-Aversion

Die Algorithmen-Aversion basiert auf den konstitutiven Arbeiten von Berkeley J. Dietvorst,<sup>102</sup> Joseph P. Simmons<sup>103</sup> und Cade Massey<sup>104</sup> in 2015.<sup>105</sup> Die Weiterentwicklungen basierend auf diesen Arbeiten gehen dabei in unterschiedliche Richtungen hinsichtlich der Erklärungsversuche und des Erkenntnisinteresses. Es wird versucht, die Ergebnisse von Dietvorst et al. (2015) in der Breite, also in andere Fachbereiche zu übertragen. Dabei wird im jeweiligen Fachbereich – wie beispielsweise dem Marketing – eine spezifische Algorithmen-Aversion gezeigt.<sup>106</sup> Außerdem wird versucht, das Phänomen der Algorithmen-Aversion in der Tiefe zu verstehen – und dabei beispielsweise zu zeigen, dass die Häufigkeit der Interaktion mit einem Algorithmus die Aversion beeinflusst.<sup>107</sup> Darüber hinaus gewinnen solche Arbeiten zunehmend an Bedeutung, die versuchen, die Algorithmen-Aversion zu mindern.<sup>108</sup> Die Schwierigkeit in der Einordnung dieser Arbeit liegt in der weitläufigen Interpretation der Algorithmen-Aversion in diesen Arbeiten. Beispielsweise werden Studien durchgeführt, die keine Incentivierung der Teilnehmer beinhalten<sup>109</sup> oder es wird auf den Effekt des nicht perfekten Forecasts verzichtet und so eine Algorithmen-Appreciation gezeigt<sup>110</sup>. Dabei zeigt sich, dass die ursprünglichen Erkenntnisse teilweise (zu) weitläufig interpretiert werden, wodurch es vereinzelt zur Kritik an den Weiterentwicklungen kommt.<sup>111</sup> Um dieser latenten Verwässerung der durch Dietvorst et al. (2015) gezeigten Effekte entgegenzuwirken, wird der Fokus in der Literaturanalyse der vorliegenden Arbeit auf die Faktoren gelegt, die eine Aversion auslösen und deren Relevanz in den Ausgangsexperimenten

---

<sup>101</sup> Siehe Abschnitt 2.2.1 und 3.2.3.

<sup>102</sup> Seit 2020 Professor für Marketing an der University of Chicago Booth School of Business, Chicago, IL. Davor Ph.D. an der The Wharton School, University of Pennsylvania.

<sup>103</sup> Dorothy Silberberg Professor of Applied Statistics an The Wharton School, University of Pennsylvania.

<sup>104</sup> Professor of the Practice an The Wharton School, University of Pennsylvania.

<sup>105</sup> Vgl. Dietvorst et al. 2015.

<sup>106</sup> Vgl. Castelo et al. 2019.

<sup>107</sup> Vgl. Efendić et al. 2020; Logg et al. 2019; Prah und van Swol 2017.

<sup>108</sup> Vgl. Dietvorst et al. 2018; Filiz et al. 2021.

<sup>109</sup> Vgl. Castelo et al. 2019.

<sup>110</sup> Vgl. Logg et al. 2019.

<sup>111</sup> Siehe den Abschnitt 2.1.5.1 in diesem Kapitel zur Arbeit von Prah und van Swol 2017.

## Begriffsbestimmungen und Stand der Forschung

von Dietvorst et al. (2015) dargelegt ist. In Abschnitt 2.1.4 wird darüber hinaus erläutert, wie die einzelnen experimentellen Elemente zur Hypothesenentwicklung beitragen. Da aufgrund der Neuartigkeit der Algorithmen-Aversion nicht von einer Theorie gesprochen werden kann, werden die postulierten Zusammenhänge als theoretisches Modell bzw. theoretischer Analyserahmen betrachtet.<sup>112</sup>

Dietvorst et al. (2015) etablieren den Begriff der Algorithmen-Aversion in ihrer Arbeit in vier Experimenten.<sup>113</sup> Sie postulieren dabei keine allgemeine Algorithmen-Aversion in dem Sinne, dass Forecaster grundsätzlich Algorithmen ablehnen. Vielmehr bezieht sich die Algorithmen-Aversion auf die zunehmende Ablehnung eines Algorithmus, sobald dessen Forecasts als potenziell fehlerhaft wahrgenommen werden. Im Folgenden wird der Aufbau der Untersuchung nach Dietvorst et al. (2015) kurz skizziert, um die wesentlichen Elemente der Untersuchung herauszuarbeiten. Diese Elemente werden in Abschnitt 2.1.4 näher erläutert.

Im Experimentalsetting wird jeweils in Kontroll- und Treatmentgruppen unterschieden. Beide Gruppen haben dasselbe Ziel, indem sie einen möglichst präzisen Forecast zu Passagierdaten eines U.S.-Bundesstaats erstellen. Eine detaillierte Beschreibung der Aufgabe findet sich in Abschnitt 3.2.1. Diese *Forecastingaufgabe* kann nur *unter Unsicherheit* beantwortet werden.<sup>114</sup> Neben einer festen Vergütung für das vollständige Durchlaufen des Experiments erhalten die Teilnehmer eine *leistungsabhängige Vergütung*. Das heißt, je genauer der Forecast ausfällt, desto höher ist die leistungsabhängige Vergütung.<sup>115</sup> Die Koppelung von leistungsabhängiger Vergütung an den Forecast definiert damit das Anspruchsniveau des *Ziels*. Für den incentivierten Forecast bekommen die Teilnehmer die Möglichkeit, ihren eigenen Forecast oder den eines Algorithmus zu wählen. Das heißt, ihre Vergütung hängt entweder von ihrem eigenen Forecast oder von dem des Algorithmus ab. Der Forecast, der darüber entscheidet, wie hoch die Vergütung ausfällt, wird im Folgenden als *incentivierter Forecast* bezeichnet. Es kann innerhalb eines Experiments für jeden Probanden mehrere

---

<sup>112</sup> Um von einer vollwertigen Theorie sprechen zu können sehen Döring und Bortz 2016, S. 57 Kriterien die für die Algorithmen-Aversion noch nicht als erfüllt betrachtet werden können. Dabei fehlen beispielsweise das in sich abgeschlossene und konsistente Aussagensystem und es liegt ein zu geringer empirischer Bewährungsgrad vor.

<sup>113</sup> Im Folgenden wird auf Studie 3b fokussiert. Studie 3b hatte die größte Anzahl an Probanden und damit die größte statistische Power. Siehe Dietvorst et al. 2015, S. 119.

<sup>114</sup> Zur Forecastingaufgabe siehe Abschnitt 3.2.1.

<sup>115</sup> Das Schema, nach dem die Probanden bezahlt werden, ist in Abschnitt 3.2.2 dargestellt.

## Begriffsbestimmungen und Stand der Forschung

solcher incentivierter Forecasts geben oder, wie für die vorliegende Untersuchung definiert, nur einen.<sup>116</sup>

Der Unterschied zwischen den Gruppen, der durch das Treatment erzeugt wird, besteht darin, dass die Treatmentgruppe Übungsforecasts machen darf. In diesen Übungsforecasts erstellen sie zehnmal einen eigenen Forecast, wobei jeweils im Anschluss an jeden der zehn Forecasts auch der Forecast des Algorithmus und der wahre Wert präsentiert werden. Diese zehnmahlige Interaktion mit dem Algorithmus durch eine Rückmeldung zur Forecastinggenauigkeit des Algorithmus und zum eigenen Forecast, wird im Folgenden als *Leistungsfeedback* definiert.

Die Kontrollgruppe hat vor ihrer Wahl für den incentivierten Forecast keine Erfahrung mit der Forecastinggenauigkeit, also der Leistung, des Algorithmus. Diese Gruppe wählt bereitwillig zu 54-57 % den Algorithmus für den incentivierten Forecast. Die Sichtbarkeit der Ergebnisse des Algorithmus – das Leistungsfeedback – führt dazu, dass in der Experimentalgruppe signifikant weniger Teilnehmer ihr Incentive durch die Wahl des Algorithmus determinieren lassen. Dietvorst et al. (2015) beschreiben die Algorithmen-Aversion daher mit: „Algorithm Aversion: People Erroneously Avoid Algorithms after Seeing Them Err“<sup>117</sup>. Burton et al. (2019) grenzen die Algorithmen-Aversion ein mit: „[T]he reluctance of human forecasters to use superior but imperfect algorithms“<sup>118</sup>. Diese Definitionen beschreiben die Algorithmen-Aversion im engeren Sinne.

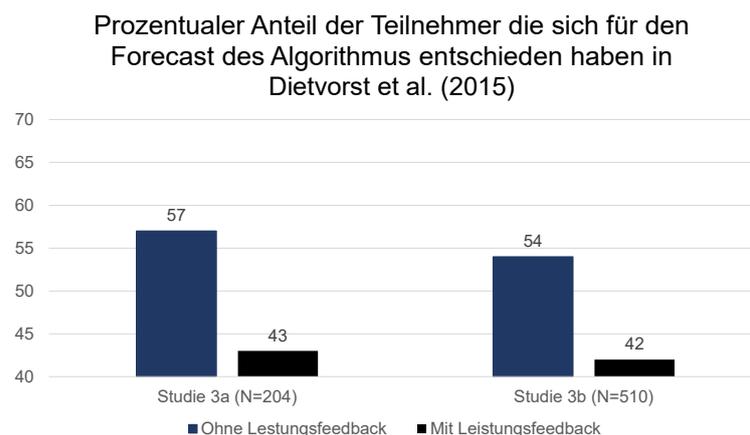


Abbildung 9: Algorithmen-Aversion nach Dietvorst et al. (2015)<sup>119</sup>

<sup>116</sup> Vergleich Studie 1 und Studie 3b in Dietvorst et al. 2015.

<sup>117</sup> Dietvorst et al. 2015, S. 114.

<sup>118</sup> Burton et al. 2019, S. 1.

<sup>119</sup> Eigene Darstellung basierend auf Dietvorst et al. 2015, S. 121.

## Begriffsbestimmungen und Stand der Forschung

Erklärt wird die Algorithmen-Aversion durch den Verlust der *Confidence* in den Algorithmus, wenn dessen Forecastinggenauigkeit wahrgenommen wird – also Leistungsfeedback gegeben wird. Zwar nehmen die Forecaster sowohl die eigene Leistung wahr als auch die des Algorithmus, jedoch führt dies lediglich zu einer Abnahme der *Confidence* in den Algorithmus und nicht der *Confidence* in den eigenen Forecast. Der Verlust der *Confidence* in den Algorithmus wiederum führt zu einer steigenden Algorithmen-Aversion. Diese Mediation ist in Abbildung 10 dargestellt. Abbildung 9 und Abbildung 10 bilden damit das für die vorliegende Arbeit zugrunde liegende theoretische Modell bzw. den theoretischen Analyserahmen.

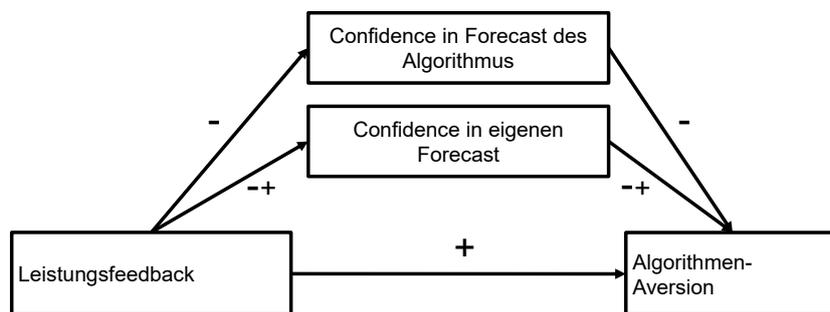


Abbildung 10: Mediation zur Erklärung der Algorithmen-Aversion<sup>120</sup>

Die Begriffsbildung in diesem Abschnitt zeigt ebenfalls, dass es wesentliche Elemente gibt, die einen besonderen Einfluss im Kontext der Algorithmen-Aversion haben. Abbildung 11 fasst diese zusammen.

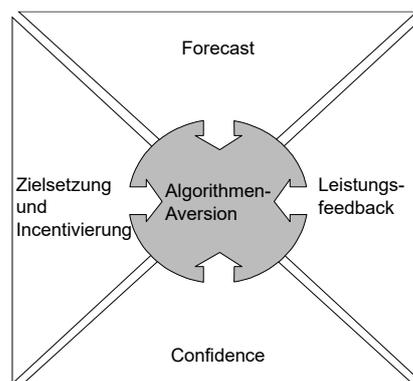


Abbildung 11: Elemente der Algorithmen-Aversion<sup>121</sup>

<sup>120</sup> Eigene Darstellung basierend auf Dietvorst et al. 2015, S. 122.

<sup>121</sup> Eigene Darstellung.

### 2.1.4 Einordnung wesentlicher Elemente der Algorithmen-Aversion in die Literatur

Die vorhergehenden Abschnitte zeigen die eindringliche Bedeutung der Algorithmen-Aversion und die Widersprüche zu anerkannten Ansichten der Agency-Theorie. Die Begriffsbildung in Abschnitt 2.1.3 zeigt ebenfalls, dass es wesentliche Elemente gibt, die einen besonderen Einfluss im Kontext der Algorithmen-Aversion haben (siehe Abbildung 11). Im Folgenden werden diese Elemente in die weiterführende Literatur eingeordnet, dadurch tiefgehend erläutert und die theoretische Fundierung wird gestärkt.

#### 2.1.4.1 Verzerrte Entscheidungen in Forecasts

Qualitativ minderwertige Daten, zu kleine Datenmengen oder schlecht trainierte Algorithmen können als technische Biases bezeichnet werden.<sup>122</sup> Diese Aspekte führen zu einem unpräzisen Forecast, der in Folge eine suboptimale Entscheidung bedingt. Die Entscheidung selbst kann dabei völlig rational und optimal getroffen werden. Neben dieser technischen Sichtweise kann der Grund für suboptimale Entscheidungen jedoch auch auf den Faktor Mensch zurückgeführt werden. Das heißt, unterschiedliche Informationen und Einflussfaktoren wirken auf den Entscheider, der daraufhin in seiner Denkweise und Entscheidungsfindung geprägt wird.<sup>123</sup> Es kann daher aus technischer Sicht eine gute Entscheidungsgrundlage vorliegen, jedoch durch unterschiedliche Charakteristika und Entscheidungsprozesse des Forecaster eine suboptimale Entscheidung getroffen werden. Abbildung 12 ordnet die Forecastingaufgabe und die damit verbundene verzerrte Entscheidungsfindung in den theoretischen Rahmen ein.

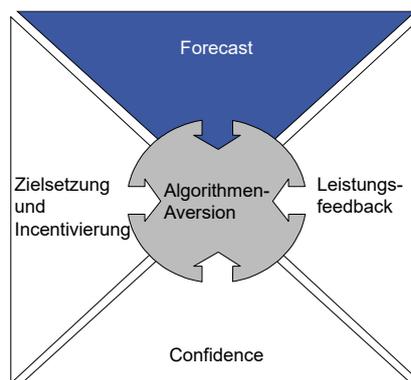


Abbildung 12: Forecasting als Element der Algorithmen-Aversion<sup>124</sup>

<sup>122</sup> Diese technisch bedingten suboptimalen Biases können auch als *Systemic Biases* bezeichnet werden. Siehe dazu Utley 2014, S. 131.

<sup>123</sup> Vgl. Golden et al. 1994, S. 33.

<sup>124</sup> Eigene Darstellung.

Die Aufgabenstellung in der Forschung zur Algorithmen-Aversion ist in dieser Situation der Entscheidungsfindung angesiedelt. In der Regel soll eine Entscheidung unter Unsicherheit, die auf einem Forecast basiert, getroffen werden. Die Literatur zur Entscheidungsfindung unter Unsicherheit erklärt grundsätzlich, wie Forecaster Entscheidungen treffen und wie die Entscheidungen durch unterschiedliche Einflussfaktoren verzerrt werden (Biases).<sup>125</sup> Dem Entscheider im Rahmen der Forschung zur Algorithmen-Aversion werden jeweils der eigene wie auch der Forecast des Algorithmus gezeigt wird. Da der Forecast des Algorithmus im Mittel deutlich besser ist, läge eine rationale Annahme darin, den Forecast des Algorithmus zu wählen und somit eine höhere Vergütung zu erreichen. Die Forecaster wählen jedoch zunehmend ihren eigenen Forecast, wodurch eine Abweichung von rationalem Handeln unterstellt werden kann. Eine Abweichung von rationalem Handeln ist in der Entscheidungsfindung aus unterschiedlichen Gründen in Forecasts nicht ungewöhnlich und bereits vielfach erforscht.<sup>126</sup> Diese Abweichungen werden als *kognitive Verzerrung* bzw. *kognitive Bias* beschrieben.<sup>127</sup>

Die Grundlagen zur Erforschung von kognitiven Biases stammen von Tversky und Kahneman (1974). Sie beschreiben das Verhalten von Individuen, die Entscheidungen unter Unsicherheit treffen. Dabei stellen sie fest, dass durch das unter Unsicherheit auftretende Risiko Forecaster ihre rationale Urteilsfindung systematisch missachten.<sup>128</sup> Unter dieser Unsicherheit ziehen Forecaster Heuristiken heran, das heißt, sie vereinfachen ihre Denkweise, um schneller zu einer vermeintlich guten Entscheidung zu gelangen. Solche Heuristiken sind in der Regel eine Vereinfachung der Entscheidung, ohne die Entscheidungsqualität zu mindern: „In general, these heuristics are quite useful, but sometimes they lead to severe and systematic errors“<sup>129</sup>. Heuristiken zur vermeintlichen Beherrschung der Entscheidung können zu systematischen kognitiven Verzerrungen, also Biases, mit negativen Konsequenzen für die Entscheidungsfindung führen. Biases verschlechtern die Qualität eines Forecasts.<sup>130</sup>

---

<sup>125</sup> Vgl. Dawes 1979, S. 571; Hutchinson et al. 2010, S. 627; Liberman et al. 2012, S. 507.

<sup>126</sup> Siehe dazu Dawes et al. 1989; Dawes 1979; Dawes und Corrigan 1974; Einhorn 1986; Fildes et al. 2009; Fildes und Beard 1992; Goodwin und Fildes 1999; Grove et al. 2000; Jacobson et al. 2011; Makridakis et al. 1982; Meehl 1954.

<sup>127</sup> Vgl. Arnott 2006, S. 59.

<sup>128</sup> Vgl. Tversky und Kahneman 1974, S. 1124.

<sup>129</sup> Tversky und Kahneman 1974, S. 1124.

<sup>130</sup> Vgl. Utley 2014, S. 131.

Je nach Entscheidungssituation und Charakteristika der Forecaster können Biases unterschiedlich stark auftreten. Es ist daher bedeutsam, zu verstehen, dass „[...] not all basic types of bias are robust across all kinds of decision processes; rather, their selective presence is contingent upon the specific processes that decision makers engage in“<sup>131</sup>. Es muss daher der zugrunde liegende Prozess, also die gesamte Entscheidungssituation, analysiert werden, um einen Bias nachhaltig erklären zu können. Im Folgenden werden die für diese Arbeit potenziell relevanten Biases aus der Literatur vorgestellt und ihr Erklärungsgehalt für die Algorithmen-Aversion aufgezeigt.<sup>132</sup>

Der bedeutendste kognitive Bias ist die *Overconfidence*.<sup>133</sup> Overconfidence tritt insbesondere in Experimentalsettings auf, ist jedoch erwiesenermaßen auch darüber hinaus ein weit verbreitetes Phänomen.<sup>134</sup> Overconfidence ist eine systematische Überschätzung der eigenen Fähigkeiten. Hribar und Yang (2016) definieren Overconfidence als „[...] an overestimation of judgmental precision or underestimation of the variance of random processes“<sup>135</sup>. Diese Selbstüberschätzung kommt häufig beim Vergleich der eigenen Meinung mit derjenigen anderer zum Ausdruck. Hier überschätzen sich die meisten Forecaster systematisch und nehmen dadurch Einbußen in der Qualität ihrer Meinung bzw. Einschätzung in hin.<sup>136</sup> Zeitgleich führt die Overconfidence zu positiv verzerrten Erwartungen gegenüber den eigenen Erfolgen bei einer Aufgabe<sup>137</sup> und dem Eingehen unverhältnismäßig großer Risiken.<sup>138</sup> Der positive Effekt der Überschätzung der eigenen Leistung liegt in einer Überschätzung des möglichen Erfolgs, wodurch mit weniger Unsicherheit agiert wird.

Die Betrachtung der Confidence in den eigenen Forecast und in den des Algorithmus ist ein wesentliches Element zur Erklärung der Algorithmen-Aversion. Daher reihen einige Forscher die Algorithmen-Aversion in die Literatur zu Overconfidence

---

<sup>131</sup> Das und Teng 1999, S. 758.

<sup>132</sup> Die hier ausgewählten und thematisierten Biases orientieren sich an Überschneidungen in Biases die in der Literatur zu Algorithmen-Aversion (Logg et al. 2019) und dem Management Accounting (Bonner 2008), Digitalisierung im Accounting (Huerta und Jensen 2017) und Forecasting (Eroglu und Croxton 2010) als relevant angesehen werden.

<sup>133</sup> Vgl. Busenitz und Barney 1997, S. 10.

<sup>134</sup> Vgl. Harvey 1997, S. 78.

<sup>135</sup> Hribar und Yang 2016, S. 221 eine ähnliche Sichtweise findet sich in Ben-David et al. 2013.

<sup>136</sup> Vgl. Logg et al. 2018, S. 100; Moore und Healy 2008, S. 502.

<sup>137</sup> Vgl. Hribar und Yang 2016, S. 226.

<sup>138</sup> Vgl. Ben-David et al. 2013, S. 1548.

ein.<sup>139</sup> Filiz et al. (2021) fassen dies zusammen mit „Algorithm aversion and overconfidence are thus presumably closely related phenomena“<sup>140</sup>. Dabei wird im Detail analysiert, ob die Einschätzung der Algorithmen-Aversion eine Ablehnung von Algorithmen ist oder schlichtweg eine Überschätzung der eigenen Meinung gegenüber der Meinung bzw. des Forecasts eines Algorithmus. Die Betrachtung der (Over-)Confidence wird in Abschnitt 2.1.4.3 detailliert vorgenommen.

Der *Overreaction Bias* besagt, dass Forecaster auf bestimmte Signale überreagieren. Sie nehmen zwar eine Anpassung ihrer Entscheidung in die richtige Richtung vor, jedoch geschieht die Anpassung auf überzogene Weise.<sup>141</sup> Eine direkte Messbarkeit dieser Überreaktion findet besonders häufig am Aktienmarkt statt, wenn neue Informationen überproportional in den Kurs einer Aktie eingepreist werden.<sup>142</sup> Eine sogenannte *Überreaktion* kann durchaus als Erklärung für eine Algorithmen-Aversion herangezogen werden. Die Information, dass ein Algorithmus keine perfekten Ergebnisse liefert, wird von Forecastern offensichtlich stark überschätzt. Da jedoch keine Anpassung der Entscheidung auf das Leistungsfeedback zum eigenen Forecast erfolgt, ist der Erklärungsgehalt durch den Overreaction Bias nicht hinreichend gegeben.

Dem Overreaction Bias entgegen steht der sogenannte *Anchoring Bias*. Dieser Bias entsteht durch das Festhalten an einem bestimmten Vergangenheitswert (Anker), auch wenn offensichtlich ist, dass eine Anpassung der Entscheidung notwendig wäre.<sup>143</sup> Es wird nur dann von ihm abgewichen, wenn zusätzlich zu den Incentives die zu erfüllende Aufgabe ausgesprochen simpel ist.<sup>144</sup> Im Umkehrschluss wirkt er in komplexen und risikoreichen Entscheidungssituationen besonders stark. Der Anchoring Bias bleibt selbst dann bestehen, wenn finanzielle Anreize gegeben werden, die bei einer Anpassung der Entscheidung größer ausfallen würden.<sup>145</sup> Die Erklärung einer Algorithmen-Aversion aus Sicht eines Ankers muss differenziert erfolgen. Ohne Leistungsfeedback besteht eine Algorithmen-Appreciation. Durch das Leistungsfeedback wird die Entscheidung zum eigenen Forecast hin verstärkt. Hier kann nicht von einem

---

<sup>139</sup> Vgl. Logg et al. 2019, S. 91.

<sup>140</sup> Filiz et al. 2021, S. 1–2.

<sup>141</sup> Vgl. Amir und Ganzach 1998, S. 333; Eroglu und Croxton 2010, S. 117.

<sup>142</sup> Vgl. Bondt und Thaler 1985, S. 89; Tuttle et al. 1997, S. 793.

<sup>143</sup> Vgl. Eroglu und Croxton 2010, S. 117 und 126.

<sup>144</sup> Vgl. Meub und Proeger 2016, S. 1.

<sup>145</sup> Vgl. Meub und Proeger 2016, S. 3.

Anchoring Bias gesprochen werden. Über die zehn Übungsforecasts im Leistungsfeedback ist ein Ankereffekt ebenfalls nicht schlüssig zu erklären, denn es wird die Entscheidung für den incentivierten Forecast getroffen, bevor der eigene potenziell incentivierte Forecast erstellt wird. Die Algorithmen-Aversion lässt sich demnach nicht durch einen Anchoring Bias erklären.

Ein weiterer einflussreicher Bias in der Erstellung von Forecasts ist der *Optimism Bias*.<sup>146</sup> Dieser beschreibt die Tendenz, Forecasts positiv verzerrt – zu optimistisch – zu erstellen bzw. zu korrigieren.<sup>147</sup> Dies kann begründet sein durch ein gewisses Wunschdenken nach einem besseren Ergebnis oder der Verknüpfung der erhöhten eigenen Wertschätzung durch Dritte mit einem positiven Forecast.<sup>148</sup> Der Optimism Bias ist demnach besonders relevant für zum Beispiel Gewinn-, Umsatz- sowie Aktienkurs-Forecasts, bei denen der Forecast einen Rückschluss auf die eigene Leistung und damit verbunden auf die Wertschätzung von Dritten darstellt.<sup>149</sup> Diese Faktoren scheinen im Themenfeld der Algorithmen-Aversion keine Relevanz zu besitzen. Die zu prognostizierenden Werte folgen keiner Zeitreihe, die bei einem Anstieg positiv wahrgenommen wird. Die Forecastingaufgabe wird daher nicht besser, indem ein Forecast möglichst hoch ausfällt, sondern, wenn dieser möglichst präzise für einen bestimmten Wert abgegeben wird.<sup>150</sup> Es hilft in den Forecastern ebenfalls nicht, einen einmal erstellten Forecast systematisch nach oben zu korrigieren. Ebenso wenig kann durch einen positiven Forecast eine Art Wertschätzung durch Dritte verbessert werden. Damit kann der Optimism Bias nicht zur Erklärung der Algorithmen-Aversion herangezogen werden.

Zusammengefasst lässt sich festhalten, dass der Overconfidence Bias teilweise eine Erklärung für die Algorithmen-Aversion darstellen kann. Die weiteren hier diskutierten Biases können lediglich argumentativ unterstützend herangezogen werden. Es wird demnach die Frage aufgeworfen, ob die Algorithmen-Aversion ein so robuster

---

<sup>146</sup> Vgl. Herrmann et al. 2008; Sun und Xu 2012.

<sup>147</sup> Vgl. Eroglu und Croxton 2010, S. 117.

<sup>148</sup> Vgl. Galanti und Vaubourg 2017, S. 326; Herrmann et al. 2008, S. 179.

<sup>149</sup> Vgl. Herrmann et al. 2008, S. 180; Sun und Xu 2012, S. 64.

<sup>150</sup> Siehe Abschnitt 3.2.1.

Mechanismus ist, dass dieser künftig selbst als *Algorithm Aversion Bias* geführt werden sollte.<sup>151</sup> Die Etablierung der Algorithmen-Aversion als Bias erfordert jedoch eine systematische und vor allem robuste kognitive Verzerrung.

### 2.1.4.2 Leistungsfeedback

Feedback ist die Konsequenz aus einer Leistung in Form der Beurteilung der erbrachten Leistung.<sup>152</sup> Durch die Möglichkeit und Funktion der Leistungsbeurteilung ist Feedback tief im Controlling verankert: „[F]eedback (information provided to a person regarding some aspect of his or her task performance) is an integral component of accounting because a fundamental role of accounting information is to facilitate individual and organizational learning“<sup>153</sup>. Feedback wird demnach gegeben, um Forecastern und Organisationen eine Weiterentwicklung zu ermöglichen. Es kann die Rückmeldung eines wahren Werts, eine konkrete Einschätzung der Leistung (98 aus 100 Punkten) oder eine einfache verbale Rückmeldung wie „gut gemacht“ für positives Feedback und „schlecht gemacht“ für negatives Feedback sein.<sup>154</sup> Die Gegenüberstellung des eigenen Forecasts mit dem des Algorithmus und dem wahren Wert gibt dem Forecaster eine Rückmeldung zur eigenen Forecastinggenauigkeit und der des Algorithmus. Die Forecastinggenauigkeit determiniert die Auszahlung und ist demnach die Forecastingleistung bzw. die *Performance*. Das dazugehörige Feedback wird als Leistungsfeedback definiert.<sup>155</sup> Abbildung 13 ordnet das Leistungsfeedback in den theoretischen Rahmen der Algorithmen-Aversion ein.

---

<sup>151</sup> Siehe Abschnitt 6.1.

<sup>152</sup> Vgl. Hattie und Timperley 2007, S. 81.

<sup>153</sup> Bonner und Sprinkle 2002, S. 329.

<sup>154</sup> Vgl. Ilgen und Davis 2000, S. 551.

<sup>155</sup> Der genaue Wortlaut ist: „After you complete the 10 practice estimates, you will go through 1 official estimate in which you will be compensated. The U.S states you see in the practice round will be different from the one you see in the official round“ Siehe Anhang 6.

## Begriffsbestimmungen und Stand der Forschung

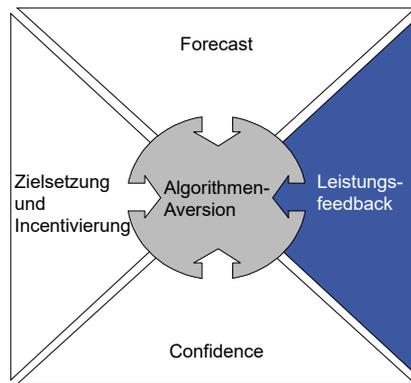


Abbildung 13: Leistungsfeedback als Element der Algorithmen-Aversion<sup>156</sup>

Diese Interpretation der Gegenüberstellung der Forecasts mit dem wahren Wert wird nicht von allen Autoren in der Literatur zur Algorithmen-Aversion vorgenommen. Berger et al. (2021) verbinden diese Rückmeldung mit dem Aufbau einer Art Vertrautheit (engl. *familiarity*): „[...] whether people’s reliance on an algorithm only decreases after becoming familiar with the algorithm, which means to observe its performance“<sup>157</sup>. Es stellt sich jedoch die Frage, ob dies eine treffende Interpretation ist. Sollte auf Vertrautheit abgezielt werden, erscheinen andere Mechanismen sinnvoller. Beispielsweise kann hier eine Beschreibung der Funktionsweise des Algorithmus oder seiner Urheber genannt werden.<sup>158</sup> Da es für die Algorithmen-Aversion jedoch relevant ist, dass der Forecaster einen fehlerhaften Algorithmus wahrnimmt und nicht *irgendeinen* Algorithmus wird in dieser Arbeit auf den Terminus des Leistungsfeedbacks zurückgegriffen.

Leistungsfeedback wurde in der Literatur zum Forecasting und der Entscheidungsfindung vereinzelt untersucht, wobei der Schwerpunkt typischerweise auf der Suche nach Ursachen und der nutzenstiftenden Wirkung von Feedback auf die Leistung liegt.<sup>159</sup> Die Literatur zeigt, dass Leistungsfeedback unterschiedliche Auswirkungen auf die Leistung der Forecaster haben kann. Sie kann die Leistung verbessern, verschlechtern<sup>160</sup> oder in bestimmten Fällen keinen Einfluss auf das Verhalten bzw. die Leistung der Forecaster nehmen.<sup>161</sup> Im Zusammenhang mit der Algorithmen-Aver-

<sup>156</sup> Eigene Darstellung.

<sup>157</sup> Berger et al. 2021, S. 56.

<sup>158</sup> Siehe Abschnitt 2.1.1.

<sup>159</sup> Vgl. Ashton 1990, S. 148; Chen et al. 2015, S. 1513.

<sup>160</sup> Vgl. Akın und Karagözoğlu 2017, S. 194.

<sup>161</sup> Eine wenig beeinflusste Entscheidung tritt insbesondere bei sehr selbstbewussten Forecastern auf. Siehe Logg et al. 2019, S. 101.

sion gibt es ein klares Bild davon, wie sich Leistungsfeedback auf die Entscheidungsfindung auswirkt: Die Wirkung von Feedback wird nicht auf die Genauigkeit des Forecasters hin untersucht, sondern vielmehr auf die Wahrscheinlichkeit hin, dass der Forecaster einem Algorithmus oder vielmehr sich selbst vertraut. Es wird postuliert, dass Forecaster ihre eigenen Forecasts bevorzugen, wenn sie Leistungsfeedback sowohl zu ihrer eigenen als auch zur überlegenen Leistung eines Algorithmus erhalten. Daher ist die Leistungsrückmeldung über die Vorhersagegenauigkeit des Algorithmus die Ursache für die Algorithmen-Aversion.<sup>162</sup> Dies legt den Fokus auf die negativen Einflüsse des Feedbacks, wodurch bestehende Ansätze weniger aussagekräftig sind. Im Folgenden werden Arbeiten vorgestellt, die sich mit Feedback in einem ähnlichen Setting wie dem der Algorithmen-Aversion befassen. Das heißt, es wird eine Wirkung von Feedback auf die Leistung untersucht, wenn Incentives als Anreizmechanismus gegeben werden.

Kluger und DeNisi (1996) resümieren in einer Meta-Analyse, dass ein Großteil der Forschungen zum Einfluss von Feedback auf die Leistung einen positiven Effekt notiert und rund ein Drittel einen negativen Einfluss. Die bedingte Wirkung von Feedback wird auf unterschiedliche Einflussfaktoren zurückgeführt. Der positive Effekt von Feedback kann beispielsweise davon abhängen, wie regelmäßig dieses gegeben wird. Erfolgt kontinuierliches Feedback, werden die Motivation und damit auch die Leistung des Einzelnen erhöht. Findet das Feedback beispielsweise sporadisch statt, kehrt sich der Effekt um und das Feedback wirkt negativ auf die Leistung.<sup>163</sup> Eine Indikation für die Algorithmen-Aversion lässt sich insofern ableiten, dass das eigene Feedback (wie auch das zum Algorithmus) regelmäßig gegeben wird und somit in der Wahrnehmung von Forecastern eine Verbesserung der Leistung erwartet wird.

Eine nähere Analyse der positiven Eigenschaften des Feedbacks führten Chew und Houghton (1996) unter Studenten durch. Leistungsfeedback fördert hier einen Lerneffekt und folglich die Leistung der Studenten. Voraussetzung dafür ist, dass adäquate Rahmenbedingungen, wie die zeitnahe Bereitstellung des Feedbacks, bestehen.<sup>164</sup> Thornock (2016) bestätigt diese Nebenbedingung, da bei unmittelbar auf die Leistung folgendem Feedback der Aufwand zu lernen für den Feedbackempfänger am

---

<sup>162</sup> Vgl. Dietvorst et al. 2015, S. 114.

<sup>163</sup> Vgl. Kluger und DeNisi 1996, S. 278.

<sup>164</sup> Vgl. Chew und Houghton 1996, S. 18.

geringsten ist.<sup>165</sup> Die unmittelbare Bereitstellung von Leistungsfeedback ist in der Algorithmen-Aversion gegeben. Daraus lässt sich schließen, dass Leistungsfeedback einen positiven Einfluss auf den Lerneffekt beim Feedbackempfänger hat. Dies würde nahelegen, dass die Feedbackempfänger durch den Lerneffekt eine Verbesserung der Einschätzung zu ihrem Forecast haben. Die Arbeit von Dietvorst et al. (2015) zeigt jedoch, dass es keine gesteigerte Confidence in den eigenen Forecast gibt.

Akın und Karagözoğlu (2017) zeigen, dass die durchschnittliche Leistung von Feedbackempfängern, die eine Rückmeldung zur eigenen Leistung erhalten haben, um 11 % schlechter ist als die Leistung derjenigen, die kein Feedback erhalten. Sie begründen dies damit, dass die Feedbackempfänger, die in ihrer Aufgabenerfüllung bereits unter hohem Konzentrationsaufwand stehen, begrenzte kognitive Reserven für das erhaltene Feedback haben.<sup>166</sup> Sie können das Feedback nicht richtig einordnen und daraus die entsprechenden Rückschlüsse zur Verbesserung ihrer Leistung ziehen. Eine Überforderung der Feedbackempfänger als Erklärung für die Algorithmen-Aversion erscheint wenig plausibel. Das Leistungsfeedback wird auf das Wesentliche begrenzt dargestellt und es gibt keine zeitliche Begrenzung zur Beurteilung der präsentierten Daten.

Eriksson et al. (2009), untersuchen in einem Laborexperiment unter anderem, wie viel Feedback, also in welcher Frequenz, bereitgestellt werden sollte, um die Leistungsfähigkeit der Feedbackempfänger zu verbessern. Als Folge eines kontinuierlichen Feedbacks werden Angst und Stress der Feedbackempfänger identifiziert, wodurch die individuelle Leistung sinkt.<sup>167</sup> In der Algorithmen-Aversion wird Feedback kontinuierlich gegeben. Dieses Verhalten steht daher tendenziell im Widerspruch zur Algorithmen-Aversion, denn der Algorithmus würde mit einer guten Leistung diesem Stress entgegenwirken können.

Lourenco et al. (2018) untersuchen den Einfluss auf die Performance durch Feedback im Kontext negativer Incentives. Während positive Incentives einen Erfolg belohnen, wird durch negative Incentives ein Misserfolg bestraft. Werden negative Incentives induziert, wirkt sich Feedback zumeist negativ auf die Leistung aus. Einen wesentlichen Einfluss auf diese Ergebnisse haben jedoch die Selbsteinschätzung und

---

<sup>165</sup> Vgl. Thornock 2016, 2-5.

<sup>166</sup> Vgl. Akın und Karagözoğlu 2017, S. 210.

<sup>167</sup> Vgl. Eriksson et al. 2009, S. 679.

die Leistung der Feedbackempfänger. Je besser die Leistung und je größer die Überschätzung der eigenen Leistung sind, desto geringer ist die Wirkung des Feedbacks. Forecaster, die ihre Leistung weniger stark überschätzen bzw. unterschätzen, reagieren stärker auf den negativen Einfluss. Besteht demnach eine gewisse Unsicherheit in der Selbsteinschätzung der eigenen Leistung, kann Feedback die Leistungsverbesserung hemmen und somit negative Effekte haben.<sup>168</sup> Eine Erklärung der Algorithmen-Aversion ist bislang ebenfalls die Überschätzung der Teilnehmer im eigenen Forecast. Weiter gibt es keine Untersuchung, die den Effekt der Leistung der Teilnehmer im Kontext der Algorithmen-Aversion untersucht.<sup>169</sup>

### 2.1.4.3 Confidence

Es wird im Folgenden zunächst eine Abgrenzung der Begriffe *Trust (Vertrauen)* und *Confidence* für diese Arbeit vorgenommen. Im Anschluss wird die Relevanz der Betrachtung von Confidence und der häufig damit einhergehenden Overconfidence vertieft. Abbildung 14 zeigt die Übersicht zu den Elementen der Algorithmen-Aversion und ordnet die Confidence hier ein.

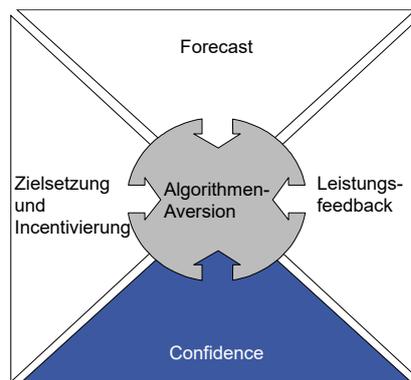


Abbildung 14: Confidence als Element der Algorithmen-Aversion<sup>170</sup>

Vertrauen kann eine wesentliche Rolle einnehmen, wenn es um die Bereitschaft von Forecastern geht, sich in unsicheren Situationen auf Automatisierungstechnologien zu verlassen.<sup>171</sup> Wesentliche Arbeiten und Theorien nutzen zur Erklärung der Interaktion mit bzw. Ablehnung von Automatisierungstechnologien den Terminus *Vertrauen*.<sup>172</sup> Aufbauend auf die Abgrenzung zum TAM (siehe Abschnitt 2.1.3) gilt es, das Verständnis von Vertrauen und Confidence in der Algorithmen-Aversion eindeutig zu

<sup>168</sup> Vgl. Lourenco et al. 2018, S. 12–14.

<sup>169</sup> Siehe hierzu Studie 2.

<sup>170</sup> Eigene Darstellung.

<sup>171</sup> Vgl. Hoff und Bashir 2015, S. 407.

<sup>172</sup> Vgl. Lee und See 2004, S. 50.

bestimmen, denn im Kontext der Algorithmen-Aversion werden diese Termini bisher nicht trennscharf unterschieden, sondern synonym verwendet:

- „[P]eople more quickly lose confidence in algorithmic than human Forecasters after seeing them make the same mistake“<sup>173</sup>.
- „[A]lgorithms are trusted [...] less for tasks that seem subjective (vs. objective) in nature“<sup>174</sup>.
- „These literatures therefore raise the question of whether individuals insufficiently trust algorithms (relative to human advisors) or merely overly trust themselves“<sup>175</sup>.

In der Literatur wird der Begriff *Vertrauen* nicht einheitlich verwendet. Dies erstreckt sich von der Uneinheitlichkeit der Definition von Vertrauen über die unterschiedliche Verwendung von Faktoren, die Vertrauen bilden, bis hin zu den Auswirkungen, die aus stark oder weniger stark ausgeprägtem Vertrauen resultieren. Vertrauen kann nach Rousseau et al. (1998) allgemein beschrieben werden als „[...] a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another“<sup>176</sup>. Vertrauen setzt demnach auf einer emotionalen Ebene die Akzeptanz für Verletzbarkeit voraus. Weiter basiert es auf positiven Erwartungen zur Einstellung des Gegenübers bzw. des Objekts, dem Vertrauen entgegengebracht wird.

Eine Einordnung zu möglichen Faktoren, die zum Aufbau von Vertrauen führen, geben Child und Möllering (2003). Sie führen beispielsweise an, dass eine individuelle Veranlagung als Teil der Persönlichkeit für die Entstehung von Vertrauen maßgebend ist. Weiter sehen sie eine erfolgreiche Zusammenarbeit in der Vergangenheit als Basis für den Aufbau von Vertrauen.<sup>177</sup> Für Deutsch (1958) ist Vertrauen neben einer Persönlichkeitsfrage und einem Konstrukt, das aus der Vergangenheit resultiert, vor allem eine rationale Entscheidung des Vertrauensgebers, die situative aktuelle Signale berücksichtigt und diese in das Vertrauen kalkuliert.<sup>178</sup> Das heißt, Vertrauen wird zusätzlich durch subjektive Wahrscheinlichkeiten bestimmt. Eine Weiterentwicklung

---

<sup>173</sup> Dietvorst et al. 2015, S. 114.

<sup>174</sup> Castelo et al. 2019, S. 809.

<sup>175</sup> Logg et al. 2019, S. 91.

<sup>176</sup> Rousseau et al. 1998, S. 395.

<sup>177</sup> Vgl. Child und Möllering 2003, S. 69.

<sup>178</sup> Vgl. Deutsch 1958, S. 265.

dazu stammt von McAllister (1995), indem er Vertrauen breit fasst und in kognitive und affektive Elemente unterteilt. Um demnach Vertrauen zu untersuchen, bedarf es der Kenntnisse über die Persönlichkeit, die Vergangenheit der Person und deren Interaktionen mit dem Objekt, dem Vertrauen entgegengebracht werden soll, und die jeweils situative Einstellung.<sup>179</sup>

Erschwert wird diese bereits komplexe Betrachtung von Vertrauen durch die umfassende Verwendung des Begriffs mit Hinblick auf das Objekt, dem Vertrauen entgegengebracht wird. Hier kann vielfältig unterschieden werden zwischen dem Vertrauen in andere Menschen, autonomes Fahren oder einfache Wettervorhersagen.<sup>180</sup> Luhmann (2000) verdeutlicht die Komplexität, die dem Vertrauenskonstrukt innewohnt, indem er Vertrauen und Confidence hinsichtlich des Objekts abgrenzt: „Trust remains vital in interpersonal relations, but participation in functional systems like the economy or politics is no longer a matter of personal relations. It requires confidence, but not trust“<sup>181</sup>. Wenn die künftige Zusammenarbeit mit Algorithmen innerhalb des Wirtschaftssystems betrachtet wird, sollte daher im Kontext der Algorithmen-Aversion von Confidence und nicht von Trust gesprochen werden. Durch den dem Vertrauen vorgelegten Charakter von Confidence kann dies als Messgröße sensibler reagieren.<sup>182</sup> Dieser Argumentation folgend erklärt sich die sinnvolle Betrachtung der Confidence in Automatisierungstechnologien, denn für eine Zusammenarbeit mit einem Algorithmus reicht diese Zuversicht aus und es ist nicht notwendig, unmittelbar Vertrauen aufzubauen.<sup>183</sup>

In Bezug auf die Arbeiten zur Algorithmen-Aversion ist damit festzuhalten, dass es bereits Erkenntnisse für ein differenziertes Bild zur Confidence in den Algorithmus und zur Confidence in den eigenen Forecast gibt. Tabelle 1 zeigt das Niveau der Confidence mit und ohne Leistungsfeedback, abgefragt anhand einer fünfstufigen Likert-Skala bei Dietvorst et al. (2015):<sup>184</sup>

---

<sup>179</sup> Vgl. McAllister 1995, S. 24.

<sup>180</sup> Vgl. Rousseau et al. 1998, S. 398.

<sup>181</sup> Luhmann 2000, S. 102. Child und Möllering 2003, S. 69 sprechen hier auch von einer „contextual confidence in institutions for building trust“.

<sup>182</sup> Vgl. Luhmann 2000, S. 102.

<sup>183</sup> Für die weiterführende Diskussion zu Vertrauen siehe auch Bohn 2007.

<sup>184</sup> Vgl. Dietvorst et al. 2015, S. 122.

## Begriffsbestimmungen und Stand der Forschung

Tabelle 1: Confidencewerte nach Dietvorst et al. (2015) (M und SD)

	Ohne Leistungsfeedback	Mit Leistungsfeedback
<b>Confidence Forecaster</b>	2,92 (0,85)	2,90 (0,80)
<b>Confidence Algorithmus</b>	3,75 (0,75)	3,36 (0,69)

Aus Tabelle 1 geht hervor, dass das Niveau der Confidence in den Algorithmus durch Leistungsfeedback abnimmt. Für die Confidence in den eigenen Forecast gilt dies jedoch nicht. Diese einseitige Abnahme der Confidence hat eine wesentliche Bedeutung (siehe Abschnitt 2.1.3): Die Erklärung für Algorithmen-Aversion wird als Beziehung zwischen dem Leistungsfeedback als unabhängiger Variable, der Confidence in den Algorithmus als Mediator und der Wahl zwischen dem eigenen und dem Forecast des Algorithmus als abhängiger Variable beschrieben.<sup>185</sup> Ein Forecaster, der Leistungsfeedback zu seiner eigenen Leistung und zu der eines Algorithmus erhält, verliert die Confidence in den Algorithmus – jedoch verliert er nicht die Confidence in seine eigene Leistung, selbst wenn diese schlechter ist als die des Algorithmus. Die Confidence in den Algorithmus wirkt als Mediator zwischen dem Leistungsfeedback und der Entscheidung des Forecasters. Die Confidence in den eigenen Forecast wurde auch als Mediator getestet, hatte aber keine signifikanten Effekte.<sup>186</sup> Im Hinblick auf die in Abschnitt 2.1.3 als wesentlich deklarierte Overconfidence muss hier differenziert werden, denn es entsteht keine absolute Overconfidence, sondern eine relative gegenüber dem Algorithmus nach Leistungsfeedback. Selbst dabei bleibt das Niveau der Confidence in den Algorithmus über dem in den eigenen Forecast, jedoch ist die Reduktion stark genug, um eine Algorithmen-Aversion auszulösen. Dies erklärt die Confidence in den Algorithmus als Mediator zwischen Leistungsfeedback und dem Einsatz von Algorithmen in Forecasts.

Am Vergleich der Confidencewerte wird in der Literatur aufgrund der vorhersehbaren Overconfidence teilweise Kritik geübt.<sup>187</sup> Die Autoren argumentieren aus Sicht der Overconfidence, dass der Forecast des Algorithmus gegen die Overconfidence der Forecaster nicht bestehen kann. Ergo ist es keine Aversion gegenüber dem Algorithmus, sondern eine Überschätzung der eigenen Fähigkeiten. Berger et al. (2021) stre-

<sup>185</sup> Siehe dazu auch Abbildung 9.

<sup>186</sup> Vgl. Dietvorst et al. 2015, S. 122.

<sup>187</sup> Vgl. Berger et al. 2021, S. 58; Logg et al. 2019, S. 91.

ben daher den Vergleich der Einschätzung der Leistung des Algorithmus mit der anderer Forecaster an.<sup>188</sup> Hierdurch wird zwar die Overconfidence reduziert, jedoch die Aufgabenstellung grundlegend verzerrt. Diese Verzerrung ist keine Möglichkeit, wie die Algorithmen-Aversion adressiert werden sollte, denn es kann in der Praxis die eigene Meinung in der Entscheidungsfindung nicht ausgeklammert und durch die Meinung eines Kollegen ersetzt werden.

### 2.1.4.4 Zielsetzung und Incentivierung

Die in Abschnitt 2.1.4.1 thematisierten Forecastingaufgaben haben einen möglichst präzisen Forecast zum Ziel. Ziele werden am effektivsten erreicht, wenn sie mit Incentives verknüpft werden.<sup>189</sup> Es gilt daher im Hinblick auf die Forecastingaufgaben: Je genauer der Forecast ausfällt, desto höher ist die Entlohnung der Forecaster. Die positive Wirkung von Incentives wurde in der jüngeren Literatur vielfach nachgewiesen.<sup>190</sup> Monetäre Incentives sind daher ein in der unternehmerischen Praxis weit verbreitetes Instrument. Des Weiteren stellen sie eine etablierte Möglichkeit dar, die Teilnehmer von experimentellen Studien zu motivieren und ihre Handlungen in eine intendierte Richtung zu lenken.<sup>191</sup> Abbildung 15 ordnet die Zielsetzung und Incentivierung in die Elemente der Algorithmen-Aversion ein.

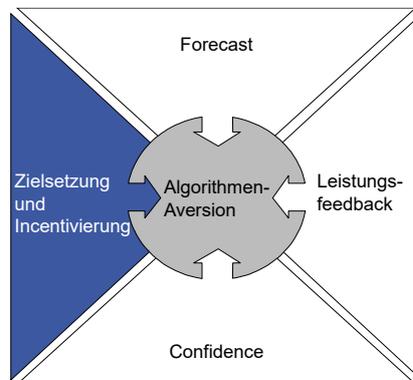


Abbildung 15: Zielsetzung und Incentivierung als Element der Algorithmen-Aversion<sup>192</sup>

In den 1960er Jahren gingen die beiden Arbeitspsychologen Locke und Latham der fundamentalen Frage nach, weshalb einige Menschen bestimmte Aufgaben besser

<sup>188</sup> Vgl. Berger et al. 2021, S. 58.

<sup>189</sup> Vgl. Corgnet et al. 2015, S. 2926. Dazu auch Bonner und Sprinkle 2002, S. 303: „Monetary incentives frequently are suggested as a method for motivating and improving the performance of persons who use and are affected by accounting information“.

<sup>190</sup> Siehe beispielsweise Garbers und Konradt 2014, S. 102 oder Shaw und Gupta 2015, S. 281.

<sup>191</sup> Vgl. Bonner und Sprinkle 2002, S. 303

<sup>192</sup> Eigene Darstellung.

erfüllen als andere und entwickelten die *Goal-Setting-Theory*.<sup>193</sup> Der Frage zu Grunde liegt die Annahme, dass menschliches Handeln sowohl durch kognitive Fähigkeiten als auch durch motivierende Faktoren bestimmt wird.<sup>194</sup> Zu dieser Zeit vertrat eine Vielzahl der Psychologen den Standpunkt, dass motivierende Faktoren ein externer Anreiz in Form von Bestätigung oder Bestrafung sind.<sup>195</sup> Diese Ansicht ergänzend gingen Locke und Latham (2013) der Frage nach den internen Faktoren nach, die die Motivation beeinflussen. Diese erhöhte Motivation wiederum soll die Leistungsfähigkeit steigern. Faktoren, die die Motivation treiben, sind hier jedoch nicht nur die unterschiedlichen Fähigkeiten oder Wissensstände, sondern vielmehr die unterschiedlichen Ziele, die Forecaster zu verschiedenen Ergebnissen führen.<sup>196</sup> Unter dem Terminus *Ziel* verstehen Locke und Latham (2013) „[...] the object or aim of an action“<sup>197</sup>.

Die beiden zentralen Aussagen der Goal-Setting-Theory sind:

- Je höher ein Ziel gesetzt wird, desto höher ist die Leistung eines Forecasters.
- Je spezifischer ein Ziel formuliert wird, desto höher ist Leistung eines Forecasters.

Die Goal-Setting-Theory ist eine der meistzitierten Theorien in der Managementforschung und hat damit eine hohe interne und externe Validität erreicht<sup>198</sup>, denn allein bis 1990 wurde die Wirksamkeit in mehr als 88 unterschiedlichen Aufgaben mit mehr als 40 000 Teilnehmern in Asien, Australien, Europa und Nord Amerika nachgewiesen.<sup>199</sup> Die Effekte wurden sowohl in Laborexperimenten als auch Feldstudien belegt und dabei mit unterschiedlichen abhängigen Variablen untermauert. Der positive Zusammenhang von Zielen auf die Leistung wurde auch in der Controlling-Forschung mehrfach dokumentiert.<sup>200</sup> Zieleeffekte sind robust auf der Ebene des Individuums, von Gruppen und ganzen Organisationseinheiten. Relevant ist außerdem die Differenzierung der Quelle des Ziels. Ziele wirken sowohl, wenn sie extern vorgegeben werden,

---

<sup>193</sup> Vgl. Latham und Locke 2013; Locke und Latham 2002, 1990.

<sup>194</sup> Vgl. Locke und Latham 1990, S. 1.

<sup>195</sup> Vgl. lock Locke und Latham 2002, S. 705.

<sup>196</sup> Vgl. Locke und Latham 1990, S. 2.

<sup>197</sup> Locke und Latham 2013, S. 4.

<sup>198</sup> Vgl. Locke und Latham 2006, S. 265.

<sup>199</sup> Vgl. Locke und Latham 2006, S. 265.

<sup>200</sup> Siehe hierzu beispielsweise Gopalakrishnan et al. 2015; Hirst und Lowy 1990; Hirst und Yetton 1999; Marginson und Ogden 2005.

als auch, wenn sie partizipativ oder in Eigenregie gesetzt werden. Zur Weiterentwicklung der Goal-Setting-Theory fordern Locke und Latham (2006): „Goal theory is an ‘open’ theory in that new elements are added as new discoveries are made“<sup>201</sup>.

Die Zielschwierigkeit wirkt, wie bereits mehrfach empirisch gezeigt, auf die Motivation von Menschen und dadurch auf deren Leistung.<sup>202</sup> Dabei gilt, dass ein schwierigeres Ziel eine höhere Motivation bedingt, dieses zu erreichen.<sup>203</sup> Dieser Effekt hält jedoch nur bis zu einem gewissen Punkt an. Voraussetzung dafür ist, dass die natürliche Kapazitätsgrenze der Menschen nicht überschritten wird, da sich die Arbeitsleistung ansonsten aufgrund von Frustration und Stress reduziert.<sup>204</sup> Von da an sind die Ziele zu anspruchsvoll und die Motivation sinkt.<sup>205</sup> Feichter et al. (2018) halten daher fest: „While targets are considered to be generally useful, their motivational effects have been shown to be conditional on the perceived difficulty of achieving them“<sup>206</sup>.

Die in Metastudien gemessenen Effektstärken der Zielschwierigkeit auf die Leistung liegen zwischen 0,52 und 0,82.<sup>207</sup> Darauf aufbauend untersuchen Locke und Latham die Effektstärken von spezifischen und schwierigen Zielen gegenüber der Anweisung, ein Ziel nach dem Kredo „gib dein Bestes“ („do your best“) zu erfüllen. Die Effektstärken in dieser Konstellation betragen 0,42 bis 0,8.<sup>208</sup> Die Kernaussage dahinter ist, dass Forecaster nicht ihr Bestes leisten, wenn ihnen gesagt wird, sie sollen ihr Bestes leisten – sie benötigen spezifische und schwierige Ziele. Die Erklärung dafür liegt in der Referenz des gesetzten Ziels. Ein Ziel, dem „do your best“ vorgegeben wird, ist vage und lässt eine Vielzahl möglicher Optimalzustände zu.

In den Studien zur Algorithmen-Aversion werden monetäre Incentives eingesetzt, um die Studienteilnehmer zu bezahlen und ihr Ziel zu definieren. Die Entlohnung besteht zumeist aus einem fixen Teil und einer leistungsabhängigen Komponente. Das heißt, die leistungsabhängige Komponente motiviert die Teilnehmer der Studie dazu, einen möglichst genauen Forecast zu erreichen. Es entsteht demnach eine Beziehung,

---

<sup>201</sup> Locke und Latham 2006, S. 265.

<sup>202</sup> Vgl. Arnold und Artz 2015, S. 61; Hollenbeck und Williams 1987, S. 204; Webb et al. 2010, S. 209.

<sup>203</sup> Vgl. Latham und Locke 1991; Locke und Latham 2002; Merchant und van der Stede 2017. Dieser Effekt ist robust für sowohl einzelne Forecaster als auch Teams (siehe Feichter et al. 2018, S. 31).

<sup>204</sup> Vgl. Waldforst 2008, S. 19.

<sup>205</sup> Siehe dazu beispielsweise Dai et al. 2018, S. 1667.

<sup>206</sup> Feichter et al. 2018, S. 31.

<sup>207</sup> Vgl. Locke und Latham 1990, S. 29.

<sup>208</sup> Vgl. Locke und Latham 1990, S. 30.

in der die Anstrengung der Teilnehmer, eine bestmögliche Leistung zu erbringen, durch ein monetäres Incentive erzeugt wird. Die motivierende Wirkung, die beste Leistung zu erbringen, führt dazu, dass Forecaster ihren Fokus auf den bestmöglichen Forecast richten. Schlussendlich tritt eine Algorithmen-Aversion auf, die dieses Verhalten negativ beeinflusst.

Ein weiterer Aspekt, den Incentives unterstützen, ist die Entwicklung von Strategien bzw. Handlungsalternativen zur Zielerreichung. Hoch gesteckte Ziele führen zu besserer Leistung als moderate, einfache oder vage Ziele.<sup>209</sup> Ziele fokussieren die Aufmerksamkeit, Anstrengungen und Handlungen des Akteurs auf das Ziel und weg von ineffektiven Handlungen.<sup>210</sup> Die Effekte von Zielen auf die Leistung sind auch davon abhängig, mit welchen Fähigkeiten ein Akteur an eine Aufgabe herantritt.<sup>211</sup> Ziele können dabei schlichtweg dazu motivieren, das bestehende Wissen und die bestehenden Fähigkeiten zielgerichtet zu *aktivieren*. Ziele können jedoch auch dazu führen, dass ein Akteur nach neuem Wissen und neuen Strategien für neue Lösungen sucht. Neue Lösungen werden insbesondere dann gesucht bzw. erzeugt, wenn eine Aufgabe unbekannt und komplex ist.<sup>212</sup> Im Falle der Algorithmen-Aversion wirken hier zwei Effekte. Zunächst werden die Teilnehmer dazu motiviert, eine Strategie zum besten Forecast zu erarbeiten. Dies kann beispielsweise dadurch erreicht werden, dass die unterschiedliche Relevanz von Daten erkannt und entsprechend gewichtet wird. Weiter ist damit die Strategie zur alternativen Ergebniserzielung mittels Algorithmenforecast gemeint.<sup>213</sup>

### 2.1.5 Stand der Forschung zur Algorithmen-Aversion

Die vorliegende Arbeit fokussiert die Algorithmen-Aversion im engeren Sinne. Das heißt, neben den bereits eingeführten Zusammenhängen der wesentlichen Elemente sind die unmittelbar darauf aufbauenden Arbeiten von Relevanz. Der Stand der Forschung wird daher zweitgeteilt. Zunächst wird ein prägnanter Überblick der Vorarbeiten gegeben – also der Literatur, auf der die Algorithmen-Aversion basiert. Diese Arbeiten entstanden vor dem Jahr 2015. Daraufaufgehend werden die Arbeiten detailliert

---

<sup>209</sup> Vgl. Locke und Latham 2006, S. 265.

<sup>210</sup> Vgl. Locke und Latham 2006, S. 265.

<sup>211</sup> Vgl. Locke und Latham 2006, S. 265.

<sup>212</sup> Vgl. Locke und Latham 2006, S. 265.

<sup>213</sup> Vgl. Bonner und Sprinkle 2002, S. 307.

beschrieben, deren Erkenntnisgewinn explizit auf der Algorithmen-Aversion im engeren Sinne basiert. Dieses Vorgehen wurde gewählt, da die Akzeptanz gegenüber Algorithmen bereits seit einigen Jahren untersucht wird. Auch wenn die Arbeitsweise von simplen Algorithmen sicherlich seit Jahrzehnten ähnlich geblieben ist<sup>214</sup>, kann nicht davon ausgegangen werden, dass die Ergebnisse der Literatur nach 70 Jahren noch gleichrangig zu werten sind. Die Darstellungsformen algorithmischer Ergebnisse oder die generell erhöhte Erfahrung mit Computern, Smartphones oder Tablets können zu Veränderungen in der Akzeptanz von Algorithmen führen.<sup>215</sup> Eine Fokussierung auf aktuelle Erkenntnisse ist demnach notwendig.

Die Forschung zur Akzeptanz von Algorithmen in der Entscheidungsfindung geht zurück auf die Arbeiten von Meehl (1954) und Dawes (1979). Meehl (1954) beschrieb als Erster die Abneigung gegen Algorithmen, indem er die grundsätzliche Überlegenheit statistischer Modelle gegenüber Expertenmeinungen hervorhob und betonte, dass Experten diese überlegenen Algorithmen immer noch mit irrationaler Skepsis bewerten und in der Folge ablehnen.<sup>216</sup> In ähnlicher Weise zeigt Dawes (1979), wie schlichte lineare Modelle den Erfolg von Studenten besser vorhersagen können als Experten, die die Studenten vermeintlich gut kennen – von letzteren jedoch abgelehnt werden.<sup>217</sup> Highhouse (2008) argumentiert, dass das Phänomen der Algorithmen-Aversion auf der Überzeugung beruht, dass menschliche Forecaster in der Lage sind, perfekte Ergebnisse zu erzielen, während Algorithmen dies nicht können.<sup>218</sup> Grove et al. (2000) weisen die Ablehnung von Algorithmen in einer Metaanalyse mit 136 Studien nach. Die Autoren stellen dabei fest, dass Algorithmen in den Studien lediglich eine durchschnittliche Überlegenheit von 10 % gegenüber Expertenmeinungen aufweisen.<sup>219</sup> Es wird festgehalten, dass bereits seit den 1950er Jahren Forschung zur Interaktion von Forecastern und Algorithmen und insbesondere der Akzeptanz gegenüber diesen betrieben wird.

---

<sup>214</sup> Beispielsweise einfache lineare Modelle zur Prognose von Zahlenreihen.

<sup>215</sup> Vgl. Quattrone 2016, S. 118.

<sup>216</sup> Vgl. Meehl 1954, S. 22.

<sup>217</sup> Vgl. Dawes 1979, S. 571.

<sup>218</sup> Vgl. Highhouse 2008, S. 333. Grove et al. 2000, S. 19 Diese Ansicht wird ebenfalls unterstützt von Einhorn 1986.

<sup>219</sup> Vgl. Grove et al. 2000, S. 19.

Der Quantensprung in der Systematisierung, Signifikanz und Begriffsbildung zur Algorithmen-Aversion gelang Dietvorst et al. (2015).<sup>220</sup> Im Folgenden werden die Arbeiten beschrieben, die auf das Grundlagenwerk von Dietvorst et al. (2015) aufbauen. Hier werden die Arbeiten ausgewählt, die zwischen dem 01.01.2016 und dem 31.12.2021 vollständig erschienen und einem Peer-Review-Prozess unterlaufen sind sowie direkt auf die Arbeit von Dietvorst et al. (2015) verweisen. Dieser Verweis kann über die Referenz auf die Algorithmen-Aversion in Form des Terms *Algorithm Aversion* im Abstract oder dem Titel der Arbeit erfolgen und bedarf einer direkten Zitation von Dietvorst et al. (2015). Anschließend wurden die jüngsten Artikel der Recherche herangezogen und mittels Schneeballsystem deren Literaturreviews und Literaturverzeichnisse durchsucht.<sup>221</sup> Auf diese Weise wurden 16 relevante Artikel identifiziert (siehe Anhang 1). Diese wurden um drei Artikel, auf die keine Zugriffsrechte bestehen<sup>222</sup> und um den Beitrag von Jung und Seiter (2021) bereinigt.<sup>223</sup>

### 2.1.5.1 Prahl und van Swol (2017)

Die erste Arbeit, die sich aufbauend auf Dietvorst et al. (2015) mit der Algorithmen-Aversion beschäftigt, publizierten Prahl und van Swol (2017) im „Journal of Forecasting“. Ihre Arbeit trägt den Titel „Understanding algorithm aversion: When is advice from automation discounted?“<sup>224</sup> Sie untersuchen anhand von 157 Studenten in einer Forecastingaufgabe, ob der Ratschlag eines erfahrenen Managers anders wahrgenommen wird als der eines Algorithmus. Ihre vermeintlich erste Erkenntnis ist eine Algorithmen-Appreciation ohne Leistungsfeedback und eine Algorithmen-Aversion mit Leistungsfeedback. Die zweite Erkenntnis ist, dass bei einem schlechten Feedback zu einem anderen Forecaster und zu einem Algorithmus die Aversion lediglich gegenüber dem Algorithmus auftritt. Die dritte Erkenntnis, dass die Teilnehmer der Studie empfinden, mehr mit einem anderen menschlichen Forecaster gemeinsam zu haben als mit einem Algorithmus, liefert einen Erklärungsgehalt zur Aversion. Den Teilnehmern der Studie ist weder der Algorithmus noch der Manager näher bekannt. Dies sehen Prahl und van Swol (2017) als eine Erklärung für eine Algorithmen-Aversion, wenn zwischen dem Ratschlag eines Algorithmus und einem anderen Forecaster entschieden wird. In

---

<sup>220</sup> Die Beschreibung der Erkenntnisse finden sich in Abschnitt 2.1.3.

<sup>221</sup> Hinzugefügt werden dabei: Efendić et al. 2020; Logg et al. 2019.

<sup>222</sup> Diese Artikel sind: Ireland 2020; Keding und Meissner 2021; Shariff et al. 2021.

<sup>223</sup> Studie 1 dieser Arbeit basiert auf Jung und Seiter 2021.

<sup>224</sup> Prahl und van Swol 2017, S. 691.

Summe sind diese Erkenntnisse keine Erweiterung der Ergebnisse von Dietvorst et al. (2015), sondern können lediglich als deren Bestätigung gesehen werden. Der fehlende Erkenntnisfortschritt könnte mehrere Ursachen haben. Zunächst interpretieren Prahl und van Swol (2017) die Arbeit von Dietvorst et al. (2015) nicht korrekt und greifen dadurch in der Interpretation ihrer empirischen Ergebnisse zu weit. Sie postulieren: „We do not replicate the finding that human advice is generally utilized more than automated advice“<sup>225</sup>. Sie übersehen daher, dass auch in Dietvorst et al. (2015) eine Algorithmen-Appreciation vorherrscht.

### 2.1.5.2 Dietvorst et al. (2018)

Dietvorst et al. (2018) zeigen durch drei Studien in „Management Science“ eine Möglichkeit zur Reduktion der Algorithmen-Aversion. Ihre Arbeit trägt den Titel: „Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them“<sup>226</sup>. In incentivierten Forecastingaufgaben hatten die Teilnehmer die Möglichkeit, die Höhe des Incentives durch ihren eigenen Forecast oder den eines Algorithmus zu bestimmen. Vorweg erhielten die Teilnehmer Leistungsfeedback, weshalb hier von einem Experiment zur Algorithmen-Aversion im engeren Sinne gesprochen werden kann. Bei Teilnehmern, die Leistungsfeedback erhalten haben, konnte die Algorithmen-Aversion signifikant gesenkt werden, wenn die Teilnehmer die Möglichkeit hatten, den Forecast des Algorithmus anzupassen. Die Arbeit greift hier ein vielfach untersuchtes Feld auf. In der Forecasting-Literatur wurde bereits mehrfach gezeigt, dass die Anpassung von Forecasts nach dem Gefühl der Forecaster weit verbreitet ist.<sup>227</sup>

Die Wahl des Forecasts des Algorithmus führte in Summe zu einer höheren Forecastinggenauigkeit und damit zu einem höheren Incentive. Grundsätzlich muss gesagt werden, dass solche Modifikationen den Forecast des Algorithmus verschlechtern. Dem entgegen Dietvorst et al. (2018) mit einer Vertiefung ihrer Untersuchung. Es ist für die Teilnehmer der Studien kaum relevant, wie stark sie das Ergebnis des Algorithmus beeinflussen können. Selbst die Möglichkeit kleinster Modifikationen reduziert die Algorithmen-Aversion. Sie resümieren daher:

---

<sup>225</sup> Prahl und van Swol 2017, S. 691.

<sup>226</sup> Dietvorst et al. 2018, S. 1155.

<sup>227</sup> Vgl. Carbone et al. 1983; Eroglu und Croxton 2010; Klassen und Flores 2001.

„[O]ur results suggest that participants' preference for modifiable algorithms was indicative of a desire for some control over the forecasting outcome, and not for a desire for greater control over the forecasting outcome, as participants' preference for modifiable algorithms was relatively insensitive to the magnitude of the modifications they were able to make“<sup>228</sup>.

### 2.1.5.3 Castelo et al. (2019)

Castelo et al. (2019) untersuchen die Algorithmen-Aversion im „Journal of Marketing Research“ und legen den Fokus auf den Charakter der Forecastingaufgabe. Ihre Arbeit trägt den Titel „Task-Dependent Algorithm Aversion“<sup>229</sup>. Der wesentliche Erkenntnisgewinn ist, dass die Algorithmen-Aversion in Aufgaben, die mit subjektiverem Charakter wahrgenommen werden, ausgeprägter ist als in Aufgaben mit objektiverem Charakter. Dabei zeigen sie, dass die wahrgenommene Objektivität einer Aufgabe messbar erhöht werden kann und dass eine solche Erhöhung der wahrgenommenen Objektivität einer Aufgabe unmittelbar die Algorithmen-Aversion senkt. Sie erklären diesen Effekt damit, dass die Teilnehmer der Studien – fälschlicherweise – einem Algorithmus die Fähigkeiten absprechen, subjektive Aufgaben erfüllen zu können. Weiter wird postuliert, dass durch die Erhöhung der wahrgenommenen Menschlichkeit eines Algorithmus die Algorithmen-Aversion auch in subjektiven Aufgaben gesenkt werden kann. Die Übertragbarkeit der Ergebnisse von Castelo et al. (2019) für die zugrundeliegende Arbeit unterliegt einigen Einschränkungen. Zunächst ist dabei zu nennen, dass die Arbeiten im Marketingkontext an Kunden durchgeführt wurden. Das heißt, es wurden beispielsweise Datinganzeigen auf Facebook geschaltet und dabei untersucht, welche Anzeigen von potenziellen Kunden häufiger angeklickt wurden. Die Anzeigen unterschieden sich dadurch, dass der *dating advice* von einem Forecaster oder von einem Algorithmus kam (subjektive Aufgabe). Analog dazu wurden Anzeigen geschaltet, in denen ein Finanzratschlag gegeben wurde (objektive Aufgabe).<sup>230</sup> Neben der grundsätzlich anderen Aufgabenstellung fehlen hier zumeist die in Dietvorst et al. (2015) eingesetzten Incentives zur Anreizsteuerung. Von Bedeutung für die vorliegende Arbeit ist damit einerseits der Nachweis der Algorithmen-Aversion i. w. S. für subjektive Aufgaben. Die Tragweite der Aversion wird damit für andere Aufgaben und

---

<sup>228</sup> Dietvorst et al. 2018, S. 1155.

<sup>229</sup> Castelo et al. 2019, S. 809.

<sup>230</sup> Vgl. Castelo et al. 2019, S. 814.

andere Probanden aufgezeigt und unterstreicht die Relevanz des Themas. Andererseits wird dargestellt, dass es im praktischen Umfeld regelmäßig auftretende Faktoren gibt, die die Aversion ganz wesentlich beeinflussen können.

Die von Castelo et al. (2019) gezeigte Reduktion der Algorithmen-Aversion durch eine Unterscheidung in subjektive und objektive Aufgabencharakteristika unterliegt in der Praxis Einschränkungen. Zwar können hierdurch Aussagen darüber getroffen werden, in welchen Fällen eine Aversion ausgeprägt ist und in welchen sie weniger ausgeprägt ist, es kann jedoch kein effektiver Eingriff in bestehende Aufgaben genommen werden. Eine subjektive Aufgabe objektiver zu charakterisieren, kann dazu führen, dass weniger – womöglich relevante – Informationen in die Aufgabe einfließen und so der Forecast verschlechtert wird. Es wird daher im Folgenden nicht als probates Mittel zur Senkung einer Algorithmen-Aversion gesehen – gleichwohl als wertvolles Wissen über das aufgabenabhängige Auftreten der Algorithmen-Aversion.

### 2.1.5.4 Logg et al. (2019)

Logg et al. (2019) fokussieren ihre Forschung zur Algorithmen-Aversion im „Organizational Behavior and Human Decision Processes Journal“ auf die sogenannte Algorithmen-Appreciation. Ihre Arbeit trägt den Titel: „Algorithm appreciation: People prefer algorithmic to human judgment“<sup>231</sup>. Die Algorithmen-Appreciation beschreibt eine generelle Präferenz für einen Algorithmus gegenüber der eigenen Einschätzung (oder der eines anderen Forecasters) – sofern es kein Leistungsfeedback zum Algorithmus gibt. Logg et al. (2019) betrachten demnach zwar die Algorithmen-Aversion, verzichten jedoch auf den Effekt, der die Algorithmen-Aversion auslöst. Sie bestätigen damit in acht Experimenten, dass ohne Leistungsfeedback zum Algorithmus eine große Präferenz für Algorithmen vorherrscht. Sie vergleichen dabei, ob ein Ratschlag von einem Algorithmus oder von einem Experten kommt.

Die Kernaussagen der Algorithmen-Appreciation nach Logg et al. (2019) ist, dass Ratschläge stärker gewichtet werden, wenn sie von einem Algorithmus kommen, als wenn sie von einem Experten stammen. Die Präferenz für den Algorithmus bleibt bestehen sowohl, wenn die Ratschläge gemeinsam, als auch, wenn sie getrennt erhalten werden. Es gab dabei zwei Fälle, in denen die Präferenz für den Algorithmus reduziert wurde: Erstens, wenn dem Algorithmus nicht ein Experte, sondern die eigene

---

<sup>231</sup> Logg et al. 2019, S. 90.

Meinung der Teilnehmer gegenübergestellt wird. Zweitens wird die Präferenz für die Gewichtung der Einschätzung eines Algorithmus dann reduziert, wenn der Teilnehmer einer Studie ein Experte im zu Grunde liegenden Aufgabengebiet ist. Das heißt, Experten gewichten ihre Meinung höher als Laien. Beide Erkenntnisse werden im engen Zusammenhang mit dem Overconfidence Bias gesehen. Der wesentliche Beitrag von Logg et al. (2019) besteht in der Definition und dedizierten Betrachtung der Algorithmen-Appreciation. Auch wenn diese bereits von Dietvorst et al. (2015) gezeigt worden ist, wurde sie von Logg et al. (2019) explizit als *Algorithmen-Appreciation* eingeführt.

Im Kontext der Experimente von Logg et al. (2019) wird nicht von einer Wahl für den Algorithmus, sondern von einer Gewichtung des Algorithmus gesprochen.<sup>232</sup> Dies ist in der Messung der Algorithmen-Appreciation durch das Judge-Advisor-System (JAS) begründet.<sup>233</sup> Hierbei wird gemessen, inwiefern eine Meinung beeinflusst wird. Dabei wird differenziert, wer die Meinung beeinflusst, und gemessen, wie stark dieser Einfluss ausfällt. Um dies zu messen, wird der Weight-on-Advice (WOA) erhoben: Dazu benötigt es eine erste Einschätzung zu einem Forecast. Darauf folgend wird dem Probanden die Meinung eines Experten oder eines Algorithmus gegeben und er kann seine Einstellung entsprechend anpassen. Diese Anpassung stellt den WOA dar. Ein WOA kann zwischen 0 % und 100 % variieren. 0 % entspricht damit keinerlei Anpassung der ursprünglichen Einschätzung, ein WOA von 100 % entspricht einer gänzlichen Anpassung der ersten Einschätzung an den Ratschlag und damit einer Verwerfung der ursprünglichen Einschätzung. Idealerweise sollten Forecaster die Entscheidung von einem Ratschlag mit ihrer ursprünglichen Einschätzung mitteln, also einen WOA von 50 % setzen.<sup>234</sup> In der Regel erfolgt jedoch nur ein WOA von 30–35 %.<sup>235</sup> Dieses Vorgehen ermöglicht eine feine Messung des Einflusses durch den Algorithmus. Gleichzeitig nimmt es Druck von der Entscheidung eines Forecasters. Es erschwert daher den Vergleich mit der Algorithmen-Aversion im engeren Sinne und wird in der vorliegenden Arbeit nicht angewendet.

---

<sup>232</sup> Siehe beispielsweise „[...] our participants placed more weight on algorithmic than human advice“ Logg et al. 2019, S. 98.

<sup>233</sup> Siehe dazu auch Snizek und Buckley 1995, S. 159.

<sup>234</sup> Vgl. Dawes und Corrigan 1974, S. 100.

<sup>235</sup> Vgl. Liberman et al. 2012, S. 511; Soll und Larrick 2009, S. 796. Gründe sind beispielsweise eine starke Selbstüberschätzung und dadurch zu starke Gewichtung der eigenen Meinung.

### 2.1.5.5 Burton et al. (2019)

Burton et al. (2019) führen einen umfassenden Literaturreview im „Journal of Behavioral Decision Making“ durch. Ihre Arbeit trägt den Titel „A systematic review of algorithm aversion in augmented decision making“<sup>236</sup>. Sie betrachten 61 einem Peer-Review unterzogene Artikel im Zeitraum von 1950 bis 2018, um der folgenden Forschungsfrage nachzugehen: „Why do people misuse (i. e., under- or over-utilize) algorithmically generated insights in augmented decision making?“<sup>237</sup> Als Motivation für Ihre Studie führen sie die Arbeiten von Dietvorst et al. (2015) an.<sup>238</sup> Sie ordnen die Algorithmen-Aversion in das Feld des *algorithmic decision making* ein und führen fünf Bereiche an, die es in der Betrachtung des algorithmic decision making zu berücksichtigen gilt:

- *Expectations and expertise*:<sup>239</sup> Ausgangspunkt dieses Felds ist es, dass private und berufliche Erfahrungen wesentlich die Erwartung an die Leistung eines Algorithmus prägen. Diese wiederum unterliegen zeitlichen Veränderungen, weshalb die Algorithmen-Aversion kein statisches Phänomen ist. Außerdem wird der Einsatz von Algorithmen erschwert, wenn sie von Experten eingesetzt werden sollen. Um den Einsatz von Algorithmen bei überzogenen Erwartungen oder bei Experten zu erhöhen, empfehlen Burton et al. (2019) Schulungen zu Algorithmen und statistischen Methoden.<sup>240</sup>
- *Decision autonomy*:<sup>241</sup> Es fehlt den Forecastern an einer *gefühlten Kontrolle* über die Leistung bzw. die Arbeitsweise des Algorithmus. Um der fehlenden Entscheidungskontrolle entgegenzuwirken, wird ein bereits seit geraumer Zeit etabliertes Prinzip des Human-in-the-Loop diskutiert. Beim Prinzip des Human-in-the-Loop wird ein Eingriff eines Forecasters in den Entscheidungsprozess oder die Arbeitsweise eines Algorithmus möglich gemacht oder gar erfordert. Für die Algorithmen-Aversion i. e. S. wurde dies bereits von Dietvorst et al. (2018) als probates Mittel gezeigt.

---

<sup>236</sup> Burton et al. 2019, S. 1.

<sup>237</sup> Burton et al. 2019, S. 1.

<sup>238</sup> Vgl. Burton et al. 2019, S. 2.

<sup>239</sup> Vgl. Burton et al. 2019, S. 4.

<sup>240</sup> Vgl. Burton et al. 2019, S. 4.

<sup>241</sup> Vgl. Burton et al. 2019, S. 4.

## Begriffsbestimmungen und Stand der Forschung

- *Incentivization*:<sup>242</sup> Bei der Incentivierung in Studien zur Algorithmen-Aversion werden widersprüchliche Argumente identifiziert. Die Einführung eines Algorithmus zur Unterstützung der Entscheidungsfindung bei einer Aufgabe erfordert zusätzliche kognitive Ressourcen beim Forecaster. Er muss neben seinem eigenen Forecast den des Algorithmus beurteilen. Er kann jedoch auch auf diese zusätzliche Beurteilung verzichten und sich weiterhin auf seine Einschätzung verlassen. Dieser zusätzliche Aufwand der Beurteilung des Algorithmus sollte nach Burton et al. (2019) durch eine höhere Incentivierung explizit berücksichtigt werden. Demgegenüber kann jedoch gesagt werden, dass durch den Algorithmus grundsätzlich ein höheres Leistungsniveau und damit auch eine höhere Vergütung erreicht werden kann.<sup>243</sup> Das bedeutet, dass durch die Korrelation von Incentive und Leistung der zusätzliche kognitive Aufwand eingepreist ist. Des Weiteren kann argumentiert werden, dass der kognitive Aufwand erheblich reduziert wird, wenn hinreichend Confidence in den Algorithmus vorherrscht und die Wahl auf den Algorithmus fällt. Ein weiterer Aspekt, der nach Burton et al. (2019) bislang keine Berücksichtigung erfährt, sind sogenannte *social incentives* als extrinsische Motivation.<sup>244</sup> In der beruflichen Praxis und im sozialen Umfeld herrschen Erwartungen, die einen Handlungsdruck zum Einsatz von Algorithmen auslösen können.
- *Cognitive compatibility*:<sup>245</sup> Das zentrale Problem, das Burton et al. (2019) hier identifiziert haben, beschreiben sie als einen Widerspruch zwischen der Entscheidung für einen Algorithmus und der eigentlichen Intuition der Forecaster. Die Arbeit mit Algorithmen im Forecasting erfordert stets zwei Denk- bzw. Entscheidungsprozesse, die zusammengeführt werden müssen: erstens die Erstellung des eigenen Forecasts und zweitens die Reflexion dieses Forecasts mit dem Forecast eines Algorithmus. Wenn dieser Abgleich auf keine befriedigende bzw. nachvollziehbare Weise möglich ist, wird der Algorithmus weniger als Hilfe gesehen, sondern vielmehr als Gefahr durch zusätzliche Unsicherheit für die richtige Entscheidung. Ein zentrales Forschungselement hier ist die Betrachtung von Algorithmen als *Black Box*. Demgegenüber sehen Burton et al. (2019)

---

<sup>242</sup> Vgl. Burton et al. 2019, S. 5.

<sup>243</sup> Vgl. Burton et al. 2019, S. 5.

<sup>244</sup> Vgl. Burton et al. 2019, S. 6.

<sup>245</sup> Vgl. Burton et al. 2019, S. 6–7.

die Intuition von Forecastern ebenfalls als eine Art Black Box, die anfällig für Verzerrungen ist.<sup>246</sup> Die Herausforderungen, um diese beiden Black Boxes zur überwinden, liegen einerseits in der Vielfalt der Algorithmen (einfache lineare Modelle vs. neuronale Netze) und andererseits in der Kontextabhängigkeit der Algorithmen-Aversion i. w. S. von Umfeld oder persönlichen Merkmalen der Forecaster.

- *Divergent rationalities*:<sup>247</sup> Hier wird hervorgehoben, dass die Erkenntnisse zur Algorithmen-Aversion häufig auf experimentelle Forschung zurückgehen, die die Praxis womöglich nicht hinreichend abbildet. Dabei wird angemerkt, dass die Entscheidungsfindung des Algorithmus auch in der Praxis beständig auf gutem Niveau liegt. Die Leistung von Forecastern hingegen kann in der Praxis je nach Aufgabe nochmals deutlich stärker variieren und unterschiedlichen Einflüssen unterliegen. Um diesem Problem zu begegnen, wird eine Fokussierung auf eine praxisnahe, realistische Entscheidungsfindungssituation angedeutet.

### 2.1.5.6 Efendić et al. (2020)

Efendić et al. (2020) fokussieren im „Organizational Behavior and Human Decision Processes“ die Reaktionszeit in der Interaktion mit Algorithmen. Ihre Arbeit trägt den Titel „Slow response times undermine trust in algorithmic (but not human) predictions“<sup>248</sup>. Sie betrachten dabei sowohl die Zeit, die ein Algorithmus für einen Forecast benötigt, als auch die Zeit, die ein Mensch für die Erstellung braucht, wobei die jeweilige Leistung eingesehen werden kann. Dabei zeigen sie mittels knapp 2000 Probanden in sieben Experimenten, dass langsame (vs. schnelle) Reaktionszeiten eines Algorithmus dazu führen, dass Forecaster dessen Genauigkeit als schlechter einschätzen und sich weniger oft auf seine Einschätzung verlassen. Die Reaktionszeit bezieht sich dabei auf die Rechendauer des Algorithmus. Dieser Effekt gilt vice versa für die Zuversicht in eine menschliche Einschätzung. Wenn hier die Reaktionszeit steigt, wird dies mit gesteigertem Arbeitseinsatz gleichgesetzt und die Entscheidung als wohlüberlegt eingeschätzt, wodurch die Zuversicht in die Einschätzung steigt. Für einen Algorithmus wird eine Aufgabe als technisch einfach betrachtet, das heißt, die Reaktionszeit hat keinen Bezug zur Qualität der Einschätzung. Eine mögliche Erklärung ist, dass die erhöhte Rechendauer des Algorithmus als *bug* oder *glitch* wahrgenommen wird.

---

<sup>246</sup> Vgl. Burton et al. 2019, S. 6. Siehe dazu auch Abschnitt 2.1.4.1.

<sup>247</sup> Vgl. Burton et al. 2019, S. 7–8.

<sup>248</sup> Efendić et al. 2020, S. 103.

Diese Untersuchungen zeigen, dass der Faktor der Zeit eine Rolle bei der Interaktion mit Algorithmen spielt.

### 2.1.5.7 Feng und Gao (2020)

Feng und Gao (2020) zeigen im „Journal Decision Support Systems“ die Algorithmen-Aversion in einem Experiment zur Bestellmengenoptimierung mittels des Zeitungsjungen-Modells. Ihre Arbeit trägt den Titel „Is optimal recommendation the best? A laboratory investigation under the newsvendor problem“<sup>249</sup>. Algorithmen werden hier als *Decision-Support-System* betrachtet. Die Forecaster müssen in den Experimenten fünfzigmal einen incentivierten Forecast durchführen. Sie sehen eine Reduktion der Algorithmen-Aversion, wenn die Teilnehmer ihrer Studie eine *Regret Aversion* entwickeln. Damit sie diese Regret Aversion zeigen können, nehmen die Teilnehmer an mehreren Runden teil und können nach jeder Runde neu entscheiden, ob sie sich für oder gegen den Algorithmus entscheiden. Dadurch, dass nicht profitmaximierende Entscheidungen bereut werden, wird in dem Fall die Wahl für die darauffolgende Runde geändert, wenn sie in der vorhergehenden Runde nicht optimal war. Hieraus sehen sie ein Spannungsfeld zwischen der Algorithmen-Aversion und einer Regret Aversion. Resultierend aus diesem Spannungsfeld schlagen sie vor, dass eine Algorithmen-Aversion dauerhaft abgebaut werden kann, wenn die Regret Aversion aktiviert wird. Das heißt, wenn der Forecaster eine bessere Prognose abgibt, wird ihm das Ergebnis des Algorithmus nicht angezeigt, wenn jedoch der Algorithmus den besseren Forecast erstellt, bekommt er diesen demonstriert.

Alle im Folgenden vorgestellten Arbeiten sind nach oder während den Datenerhebungen dieser Arbeit erschienen. Keine der Arbeiten steht im Widerspruch zu den Erkenntnissen dieser Arbeit. Sie können demnach als komplementär angesehen und argumentativ hinzugezogen werden.

### 2.1.5.8 Berger et al. (2021)

Berger et al. (2021) publizieren im „Journal für Business & Information Systems Engineering“ einen Beitrag zur Algorithmen-Aversion, der sich mit der Fähigkeit von Algorithmen zum Lernen beschäftigt. Ihre Arbeit trägt den Titel „Watch Me Improve – Algorithm Aversion and Demonstrating the Ability to Learn“<sup>250</sup>. Um lernfähige Algorithmen

---

<sup>249</sup> Feng und Gao 2020, o. S.

<sup>250</sup> Berger et al. 2021, S. 55.

anzuwenden, werden diese den Teilnehmern der Studie im Kontext des Machine Learning vorgestellt. In einem incentivierten Experimentaldesign auf Amazon Mechanical Turk (MTurk) mit 452 Teilnehmern, gilt es für die Forecaster, die in einem Callcenter eingehenden Anrufe abzuschätzen. Zu dieser Aufgabe gibt es zunächst für alle Teilnehmer acht Übungsrounds und im Anschluss eine incentivierte Einschätzung. Durch die Experimente konnte zunächst gezeigt werden, dass der Effekt des Leistungsfeedbacks die Confidence in einen Algorithmus, jedoch nicht die Confidence in einen *Human Advisor* reduziert. Eine Erklärung dafür, die Berger et al. (2021) sehen, ist, dass Forecaster ihre fehlerhaften Einschätzungen mit der Zeit ausmerzen, während Algorithmen in ihrer fehlerhaften Leistung vermeintlich *feststecken*. Um dies näher zu erforschen, präsentierten sie den Teilnehmern einen Algorithmus, der im Laufe der Übungsrounds seine Präzision verbessert. Das heißt sie vergleichen einen Algorithmus mit konstanter Präzision mit einem Algorithmus, der zunächst sehr unpräzise ist (unpräziser als der Kontrollalgorithmus) und im Laufe der Übungsrounds immer besser wird (besser als der Kontrollalgorithmus). Gemessen wird dabei, wie in den Studien von Logg et al. (2019), der WOA. Berger et al. (2021) konnten damit zeigen, dass die Teilnehmer mehr auf einen lernenden als auf einen nicht lernenden Algorithmus setzen. Des Weiteren postulieren sie, dass sie keinen Unterschied zwischen dem Vertrauen in einen lernenden Algorithmus und einen lernenden Human Advisor zeigen können. Sie leiten daraus ab, dass lernende Algorithmen die Algorithmen-Aversion senken können. Dies kann als bedeutsame Erkenntnis gesehen werden, denn in der Praxis werden Datenmengen im Laufe der Zeit anwachsen und Algorithmen mit steigenden Daten besser trainiert. Limitierend wirkt, dass die Erkenntnisse von Berger et al. (2021) eben nur auf solche lernfähigen Algorithmen angewendet werden können.

### 2.1.5.9 Renier et al. (2021)

Renier et al. (2021) fokussieren im „Journal Computers in Human Behavior“ mit ihrer Arbeit „To err is human, not algorithmic – Robust reactions to erring algorithms“<sup>251</sup> die durch Leistungsfeedback ausgelösten Reaktionen bei Forecastern. Sie bleiben dabei eng orientiert an der Arbeit von Dietvorst et al. (2015) und versuchen, das Verständnis der Algorithmen-Aversion auf einer psychologischen Ebene zu erweitern. In Online-Experimenten mit 709 Teilnehmern untersuchen sie dies einerseits auf unterschiedli-

---

<sup>251</sup> Renier et al. 2021, o. s.

che Reaktionen von Forecastern auf Leistungsfeedback zum Algorithmus und zu anderen Forecastern, und andererseits auf unterschiedliche Rahmenbedingungen zum Einsatz von Algorithmen. Die Reaktionen der Probanden werden unterschieden in *gut reactions*, *justice cognitions* und *behavioral intentions*.<sup>252</sup> Dabei durchlaufen Forecaster diese drei Stadien der Reaktionen sequenziell. Renier et al. (2021) zeigen, dass bei Algorithmen (gegenüber einem Forecaster) die *gut reactions* stärker ausgeprägt sind, das heißt, es liegt eine geringere Akzeptanz vor und es werden mehr negative Gefühle zum Ausdruck gebracht. Demgegenüber waren die *justice cognitions* weniger ausgeprägt bei Algorithmen (gegenüber einem Forecaster). Das heißt, die Haltung war von weniger Schuldzuweisung, aber auch von weniger Nachsicht oder Vergebung geprägt. Schließlich sind die *behavioral intentions* gegen Algorithmen stärker und führen zur Algorithmen-Aversion. Die Aversion gegenüber Algorithmen ist demnach von hohen Erwartungen an Algorithmen geprägt, wobei die Enttäuschung über diese hohen Erwartungen keine starke emotionale Reaktion ist. Emotionale Reaktionen wie Vergebung bleiben zwischenmenschlicher Interaktion vorbehalten.<sup>253</sup> Dies unterstreicht die Überlegungen zum Vertrauen bzw. zur Confidence in Algorithmen in Abschnitt 2.1.4.3. Der zweite Teil der Untersuchung von Renier et al. (2021) zeigt, dass sowohl die Algorithmen-Aversion als auch die drei Stadien der Reaktion robust sind. Die Robustheit zeigt sich in drei Dimensionen:

- Reifegrad des Algorithmus (Algorithmus hat eine bessere oder gleich gute Performance wie der Forecaster)
- Auswirkung der schlechten Performance des Algorithmus auf das Incentive (stark oder gering)
- Anwendungsbereich des Forecasts (Recruiting und Finance).

In Summe schärfen diese Erkenntnisse die Algorithmen-Aversion und zeigen, dass es weniger relevant ist, zu untersuchen, ob der Vergleich zwischen einem anderen Forecaster und dem Algorithmus gemacht wird, sondern, wie die robuste Algorithmen-Aversion beim Entscheider selbst wirkt und wodurch sie womöglich beeinflusst wird.

---

<sup>252</sup> Die Unterscheidung in diese drei Stufen basiert auf dem *Third-party Reactions to Mistreatment Model* nach O'reilly und Aquino 2011.

<sup>253</sup> Renier et al. 2021, o. S.

### 2.1.5.10 Niszczoła und Kaszás (2020)

Niszczoła und Kaszás (2020) publizieren im „PLOS ONE“ empirische Ergebnisse zur Algorithmen-Aversion in Investitionsentscheidungen. Ihre Arbeit trägt den Titel: „Robo-investment aversion“<sup>254</sup>. Dabei steht die These im Mittelpunkt, dass Algorithmen insbesondere dann abgelehnt werden, wenn Investitionsentscheidungen eine moralische Kontroverse enthalten. Anhand von 3828 Probanden zeigen sie, dass Algorithmen die Fähigkeit abgesprochen wird, moralisch richtige Entscheidungen zu treffen. Dabei werden die Teilnehmer in einem von fünf Experimenten für ihre Entscheidung incentiviert. Hierbei zeigt sich, dass es keine grundsätzliche Ablehnung von Algorithmen gibt. Allerdings tritt eine Aversion auf, wenn moralische kontroverse Entscheidungen getroffen werden müssen. In den Studien wird kein Leistungsfeedback gegeben. Es kann daher resümiert werden, dass, analog zu den subjektiv wahrgenommenen Aufgaben in Castelo et al. (2019), moralisch fragwürdige Aufgaben zu einer Algorithmen-Aversion führen können.<sup>255</sup> Die moralische Kontroverse muss dabei als Auslöser der Algorithmen-Aversion betrachtet werden, was die Übertragbarkeit der Ergebnisse erschwert.

### 2.1.5.11 Filiz et al. (2021)

Filiz et al. (2021) haben im „Journal of Behavioral and Experimental Finance“ gezeigt, dass Erfahrungen im Umgang mit einem Algorithmus die Algorithmen-Aversion reduzieren kann. Ihre Arbeit trägt den Titel „Reducing algorithm aversion through experience“<sup>256</sup>. Erfahrung meint im Sinne dieser Untersuchung, dass die Forscher Leistungsfeedback anhand von vierzig anstelle von zehn Übungsforecasts untersuchen. Die Arbeit von Filiz et al. (2021) erweitert damit das Verständnis der Algorithmen-Aversion insofern, dass die Algorithmen-Aversion womöglich eine temporäre Erscheinung in der Zusammenarbeit mit neuen Algorithmen ist. Die Übertragbarkeit der Erkenntnisse auf die Algorithmen-Aversion i. e. S. unterliegt jedoch einigen Einschränkungen. Zunächst ist dabei die Aufgabenstellung mit dazugehöriger Incentivierung zu nennen. Die Teilnehmer der Studie mussten den Verlauf eines Aktienkurses anhand von vier Datenpunkten vierzigmal vorhersagen. Dabei gab es stets nur die Antwortmöglichkeiten, dass der Aktienkurs steigt oder fällt. Den Teilnehmern wird mitgeteilt, dass der Algorithmus mit 70 %iger Richtigkeit arbeitet. Für jeden der 40 Forecasts erhalten sie die

---

<sup>254</sup> Niszczoła und Kaszás 2020, o. S.

<sup>255</sup> Vgl. Castelo et al. 2019, S. 809.

<sup>256</sup> Filiz et al. 2021, o. S.

Wahl zwischen dem Algorithmus und der eigenen Entscheidung und werden anschließend für jeden einzelnen Forecast entlohnt. Eine weitere Einschränkung der Übertragbarkeit der Ergebnisse liegt darin, dass der Algorithmus am theoretischen Optimum agiert. Das heißt, die Teilnehmer können mit sehr viel Zeitaufwand theoretisch das Ergebnis des Algorithmus erreichen, jedoch systematisch nicht besser als der Algorithmus sein. Eine weitere Kritik an der Arbeit von Filiz et al. (2021) ist die Fülle an Forecasts, die die Teilnehmer im Experiment durchführen.<sup>257</sup> Bei 40 Forecasts setzt neben einem Lerneffekt womöglich eine Ermüdung ein und es entsteht daraus eine Tendenz zum Algorithmus.

### **2.1.6 Zusammenfassung der bestehenden Erkenntnisse zur Algorithmen-Aversion.**

Obwohl für den Stand der Forschung im vorhergehenden Abschnitt nur solche Arbeiten ausgewählt wurden, die unmittelbaren Bezug auf die grundlegenden Erkenntnisse von Dietvorst et al. (2015) nehmen, zeigt sich, dass die Algorithmen-Aversion in der Literatur nicht einheitlich definiert und betrachtet wird. Aus der Unschärfe in der Definition wiederum entstehen Interpretationsspielräume im Forschungsdesign und der empirischen Ergebnisse, woraus teilweise verzerrte Schlüsse zur Ausprägung der Algorithmen-Aversion gezogen werden. Um dieser Unschärfe entgegenzuwirken, wird in dieser Arbeit unterschieden in Algorithmen-Aversion i. e. S. und Algorithmen-Aversion im i. w. S.

Die Algorithmen-Aversion i. e. S. ist hochgradig spezifisch.<sup>258</sup> Sie tritt auf in Forecasts, in denen Forecaster Leistungsfeedbacks zu Algorithmen und ihrer eigenen Leistung erhalten. Eine Erklärung zur Algorithmen-Aversion i. e. S ist der Verlust der Confidence in den Forecast eines Algorithmus nach Leistungsfeedback zum Forecast. Weitere wiederkehrende Elemente sind die Verknüpfung des Forecasts mit einem Incentive. Der Algorithmen-Aversion i. e. S. inhärent ist eine Algorithmen-Appreciation, bei der ohne das besagte Leistungsfeedback Algorithmen bereitwillig genutzt werden.<sup>259</sup> Wird in einem Experiment mit diesen Elementen interagiert, kann von einer Algorithmen-Aversion i. e. S. gesprochen werden. Die Literatur gibt Hinweise, dass ein

---

<sup>257</sup> Vgl. Filiz et al. 2021, o. S.

<sup>258</sup> Vgl. Dietvorst et al. 2015, S. 114.

<sup>259</sup> Vgl. Logg et al. 2019, S. 90; Prahel und van Swol 2017, S. 691.

langsamer Algorithmus die Aversion erhöht,<sup>260</sup> und dass die Anzahl der Übungsforecasts einen Einfluss auf die Algorithmen-Aversion hat.<sup>261</sup> Weiterhin wird sie abgebaut durch eine Möglichkeit, den Output eines Algorithmus anzupassen.<sup>262</sup> Analog dazu wird die Aversion reduziert, wenn der Algorithmus selbst lernfähig ist.<sup>263</sup>

Die Algorithmen-Aversion i. w. S. kann nur unscharf beschrieben werden. Durch die Unschärfe kann eine enumerative Beschreibung über den Stand bisheriger Forschung erfolgen. Die Erkenntnisse aus der Algorithmen-Aversion sind für die Betrachtung der Algorithmen-Aversion i. e. S. nicht vollständig zu vernachlässigen. Denn auch, wenn sie nicht alle Elemente der Algorithmen-Aversion i. e. S. berücksichtigt, ist es naheliegend, dass die empirisch gezeigten Zusammenhänge auch in einem Setting zur Algorithmen-Aversion i. e. S. auftreten können. Es sind hier insbesondere die Robustheit der Algorithmen-Aversion in den unterschiedlichen Aufgaben (beispielsweise Recruiting, Marketing, Finance)<sup>264</sup> zu nennen. Außerdem ist der Charakter von Aufgaben zu unterscheiden. Der Charakter meint beispielsweise die Subjektivität<sup>265</sup> einer Aufgabe oder moralische Aspekte<sup>266</sup>. Diese Erkenntnisse sind nicht in einer Linie mit früheren Thesen, dass die mangelnde Perfektion von Algorithmen eine Aversion auslöst.<sup>267</sup> Forecaster beurteilen Algorithmen technokratisch und sehen daher kontextspezifische Grenzen für die Fähigkeit von Algorithmen zu guten Forecasts.<sup>268</sup>

Die Literatur zeigt, dass die Algorithmen-Aversion kein statisches zu betrachtendes Phänomen ist. In den Untersuchungen dazu, wie sie wirkt und welchen Einflussfaktoren sie unterliegt, wird häufig eine Variation des Algorithmus oder der Aufgabe betrachtet. Dabei gibt es bisher seit Dietvorst et al. (2015) kaum Indikationen, wie die Umstände, in denen sich Forecaster bei der Erstellung eines Forecasts befinden – also deren Arbeitsumfeld – mit der Algorithmen-Aversion interagiert. Für diese Arbeit wurde das Controlling als besonders relevantes Feld der Algorithmen-Aversion

---

<sup>260</sup> Vgl. Efendić et al. 2020, S. 103.

<sup>261</sup> Vgl. Feng und Gao 2020, o. S.; Filiz et al. 2021, o. S.

<sup>262</sup> Vgl. Dietvorst et al. 2018, S. 1155.

<sup>263</sup> Vgl. Berger et al. 2021, S. 55.

<sup>264</sup> Vgl. Castelo et al. 2019, S. 809; Renier et al. 2021, o. S.

<sup>265</sup> Vgl. Castelo et al. 2019, S. 809.

<sup>266</sup> Vgl. Niszczoła und Kaszás 2020, o. S.

<sup>267</sup> Vgl. Highhouse 2008, S. 333.

<sup>268</sup> Vgl. Renier et al. 2021, o. S.

identifiziert. Wie Algorithmen hier eingesetzt werden und welche Aspekte aus dem Arbeitsumfeld besonders relevant sind, wird im Folgenden beschrieben.

### 2.2 Einordnung der Algorithmen-Aversion in das Controlling

Das *Controlling* ist in dieser Terminologie ein primär deutsches Konzept. Trotz des anglo-amerikanischen Wortursprungs besteht in der internationalen Forschung kein exaktes Äquivalent. In der Regel werden die Begriffe *Management Accounting* oder seltener *Management Control* gebraucht. In der vorliegenden Arbeit wird der Begriff *Controlling* stellvertretend verwendet.<sup>269</sup>

Anthony (1965) beschreibt das Controlling mit „[...] the process by which managers assure that resources are obtained and used effectively and efficiently in the accomplishment of the organization's objectives“<sup>270</sup>. Die Aufgabe eines Controllers besteht damit darin, eine Systematik aufzubauen, die es dem Management ermöglicht, eine effektive Steuerung und Kontrolle auszuüben.<sup>271</sup> Das Controlling schafft durch diese Systematik konsistente Beziehungen und Prozesse, welche die Planung, Steuerung und Informationsbereitstellung ermöglichen. Weiter hat das Controlling zur Aufgabe, diese Beziehungen und Prozesse zu koordinieren und die Koordination effizient zu gestalten.<sup>272</sup> Das Management zu unterstützen, bedarf der Fähigkeit, zunächst objektive Entscheidungsgrundlagen zu schaffen, die wiederum eine rationale Entscheidungsfindung ermöglichen.<sup>273</sup> Klassische Aufgaben wie Analysen, Forecasts, Optimierung und Simulation des Controllings sind daher häufig datengetrieben.<sup>274</sup> Im Controlling erweitern Algorithmen die Möglichkeiten dieser Aufgaben, da sie in der Lage sind, große und unstrukturierte Datenmengen zu erfassen und zielgerichtet auszuwerten.<sup>275</sup>

Die Durchführung dieser Aufgaben mit neuen Algorithmen erfordert auch neue Fähigkeiten von Controllern und die rationale Entscheidungsfindung in diesen Aufgaben unterliegt einer Vielzahl von Einflussfaktoren, wie den kognitiven Verzerrungen in

---

<sup>269</sup> Eine Subsumierung des Management Accountings und des Controllings ist in der deutschen Forschung gängige Praxis. Siehe beispielsweise Obermaier und Müller 2008.

<sup>270</sup> Anthony 1965, S. 27.

<sup>271</sup> Vgl. Anthony 1965, S. 28.

<sup>272</sup> Vgl. Horváth 1978, S. 202.

<sup>273</sup> Vgl. Weber und Schäffer 2001, S. 1–6.

<sup>274</sup> Vgl. Mehanna et al. 2018, S. 503–505.

<sup>275</sup> Vgl. Al-Htaybat und Alberti-Alhtaybat 2017, S. 851; Griffin und Wright 2015, S. 377

Forecasts – wie eine Algorithmen-Aversion –, dem Opportunismus oder Gewohnheiten.<sup>276</sup> Dabei wird die Informationsbereitstellung durch das Controlling immer dezentraler und unkomplizierter. Schäffer und Weber (2016) sprechen daher von einer zunehmenden Demokratisierung des Informationszugangs, welche vom Controller gefördert werden sollte.<sup>277</sup> Das heißt, das Controlling hat die Funktion des Gatekeepings zu den Daten. In Zukunft wird dieses jedoch verringert und die Daten und Auswertungen werden vermehrt geteilt, wodurch ein Kontrollverlust entstehen kann. Wenn die Funktion des Gatekeepings entfällt, ist es umso relevanter zu verstehen, in welchen Prozessen eine Algorithmen-Aversion auftritt. Fehlt dieses Verständnis, stehen den Empfängern der Forecasts zwar die Auswertungen zur Verfügung, sie werden jedoch nicht in der Entscheidungsfindung berücksichtigt.

Becker und Brandt (2016) betonen darüber hinaus die Aufgabe des Controllings in der Aufdeckung wertvernichtender Aktivitäten: „Das Controlling hat allein schon aufgrund seiner originären Funktion der Lokomotion die Aufgabe, die tatsächliche Wertschöpfung permanent zu überwachen. Dies umfasst die Identifikation von Potentialen zur Wertsteigerung und die Aufdeckung wertvernichtender Aktivitäten“<sup>278</sup>. Die Aversion gegenüber Algorithmen kann als solche wertvernichtende Aktivität verstanden werden, weshalb es eine wesentliche Aufgabe des Controllings ist, zu verstehen, wann eine Algorithmen-Aversion auftritt, um diese wiederum durch eine effektive Steuerung langfristig zu reduzieren.<sup>279</sup>

Da die Aufgabe des Controllings nicht nur ist, die Informationen für das Management bereitzustellen, sondern ebenso die Qualität der Prozesse zur Informationsaufbereitung sicherzustellen, erscheint die Algorithmen-Aversion in zweifacherweise eine für das Controlling relevante Problemstellung zu sein (siehe Abbildung 16): zum einen die Kenntnis zur Vermeidung und Steuerung der wertvernichtenden Algorithmen-Aversion und zum anderen die auftretende Algorithmen-Aversion in den Controlling-Prozessen selbst.

---

<sup>276</sup> Vgl. Obermaier und Müller 2008, S. 327–328.

<sup>277</sup> Vgl. Schäffer und Weber 2016, S. 11.

<sup>278</sup> Becker und Brandt 2016, S. 464.

<sup>279</sup> Vgl. Luft 2016, S. 9; Malmi und Brown 2008, S. 288.

## Begriffsbestimmungen und Stand der Forschung

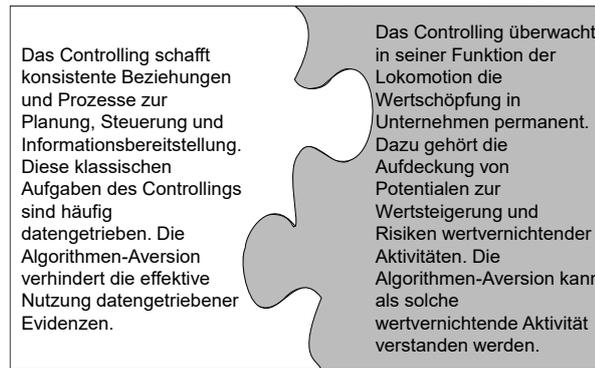


Abbildung 16: Die zweifache Bedeutung der Algorithmen-Aversion im Controlling<sup>280</sup>

### 2.2.1 Die Bedeutung von Business Analytics für Algorithmen im Controlling

Die wachsende Bedeutung von Algorithmen im Controlling liegt wesentlich in der digitalen Transformation von Unternehmen begründet.<sup>281</sup> Die digitale Transformation fördert einen ständigen Wandel der Anforderungen und Weiterentwicklung der Möglichkeiten durch Business Analytics im Controlling.<sup>282</sup> Chamoni und Gluchowski (2017) definieren Business Analytics als: „Sammlung unterschiedlicher Methoden und Techniken, welche dazu dienen, Erkenntnisse aus verfügbaren Daten für unternehmerische Entscheidungen zur Steuerung der Geschäftsprozesse zu gewinnen“<sup>283</sup>. Hinsichtlich der Definition des Controllings in Abschnitt 2.2 kann Business Analytics als integrales System des Controllings verstanden werden. Dabei verknüpft Business Analytics drei unterschiedliche Funktionen bzw. Bereiche miteinander: Informationssysteme, statistische Methoden und Algorithmen sowie das Controlling.<sup>284</sup> Ziel von Business Analytics ist vordergründig die Erhöhung des Unternehmenserfolgs durch fakten- und datenbasierte analytische Modelle in der Entscheidungsfindung.<sup>285</sup> Eben diese fakten- und datenbasierten analytischen Modelle beruhen auf Algorithmen.<sup>286</sup>

<sup>280</sup> Eigene Darstellung.

<sup>281</sup> Vgl. Holthoff und Decher 2020, S. 53; Sejdić 2020, S. 61.

<sup>282</sup> Vgl. Al-Htaybat und Alberti-Alhtaybat 2017, S. 850; Möller et al. 2020, S. 1; Quattrone 2016, S. 118.

<sup>283</sup> Chamoni und Gluchowski 2017, S. 9.

<sup>284</sup> Vgl. Schläfke et al. 2013, S. 116.

<sup>285</sup> Vgl. Mehanna et al. 2018, S. 502. Dieser Definition lässt sich entnehmen, dass Business Analytics auch in anderen betriebswirtschaftlichen Funktionen wie dem Marketing oder Human Resources Einzug hält. Siehe dazu Seiter 2019, S. 12.

<sup>286</sup> Vgl. Seiter 2019, S. 2.

## Begriffsbestimmungen und Stand der Forschung

Ermöglicht und getrieben wird die Rolle von Business Analytics durch die ständige Weiterentwicklung von ERP-Systemen. Auch wenn Daten bereits in der Vergangenheit zur Verfügung standen, gibt es zunehmend bedeutende Entwicklungen hin zu

- mehr Daten in einem immer größeren Detailgrad und<sup>287</sup>
- besserem Datenzugriff bei umfassenderer Datenverknüpfung.<sup>288</sup>

Diese Entwicklungen gehen häufig weit über finanzielle Daten hinaus und vereinen große Datenmengen aus internen und externen Quellen und Datenbanken.<sup>289</sup> Für das Controlling ist eine Business Analytics-Implementierung vor dem Hintergrund immer größer werdender Datenmengen und Analysemöglichkeiten essenziell. So besteht die Möglichkeit, komplexe Strukturen von Daten zu verarbeiten. Es können beispielsweise digital *Forecasts* konzipiert werden, wobei Strategien und Business Cases durch Algorithmen geprüft und bewertet werden.<sup>290</sup> Nebst solch intelligenter Unterstützung in *Forecasts*, kann Business Analytics das Controlling in repetitiven Regelprozessen unterstützen. Die Einführung von Robotic Process Automation beispielsweise ermöglicht es, die Vor- und Aufbereitung von Daten völlig automatisiert durchzuführen.<sup>291</sup> Ergo gibt es eine schnellere Verfügbarkeit von Auswertungen, die auf größeren Datenmengen beruhen und eine geringere Fehleranfälligkeit haben. Weiter ist es durch Algorithmen möglich, komplexe Strukturen und Muster in großen Datenmengen zu entdecken, die mit kognitiver Arbeit häufig nicht erkannt werden.<sup>292</sup>

Die folgende Systematisierung von Algorithmen im Controlling bzw. Business Analytics folgt Gluchowski (2016) und Seiter (2019). Es wird dabei unterschieden in *Beschreibungsmodelle* bzw. *Descriptive Analytics*, *Prognosemodelle* bzw. *Predictive Analytics* und *Optimierungsmodelle* bzw. *Prescriptive Analytics*. Während in der Vergangenheit ein Fokus auf der deskriptiven Auswertung lag, ist es zunehmend von Relevanz, die unterschiedlichen Daten in großen Mengen so zu vereinen, dass Aussagen

---

<sup>287</sup> Vgl. Appelbaum et al. 2017, S. 31.

<sup>288</sup> Vgl. Seiter 2019, S. 5.

<sup>289</sup> Vgl. Schläfke et al. 2013, S. 110; Seiter 2019, S. 2–3.

<sup>290</sup> Vgl. Mehanna et al. 2018, S. 502.

<sup>291</sup> Vgl. Weißenberger 2021, S. 10.

<sup>292</sup> Vgl. Kleinberg et al. 2018, S. 22; Weißenberger 2021, S. 8.

über die Zukunft getroffen werden können<sup>293</sup>, und darüber hinaus datengetriebene Modelle zur Lösungsoptimierung herangezogen werden können.<sup>294</sup>

Mittels Descriptive Analytics werden Daten zusammengefasst, um daraus Muster abzuleiten. Daten werden kategorisiert, charakterisiert und klassifiziert, um diese in nützliche Informationen umzuwandeln. Diese Daten beschreiben die Vergangenheit und Gegenwart. Deskriptive Analytics sind in der Unternehmenspraxis bereits etabliert und finden breite Anwendung beispielsweise im Management Reporting. Üblicherweise werden dafür unter anderem Clustermodelle, deskriptive Statistiken oder Text Mining eingesetzt.<sup>295</sup>

Predictive Analytics dient dazu, Datenpunkte zu identifizieren und in einen sinnvollen logischen Zusammenhang zu bringen, damit diese als Prädiktoren für künftige Datenpunkte der gleichen Datenbasis dienen. Algorithmen, die im Predictive Analytics eingesetzt werden, sind beispielsweise die Regressions-, Klassifikations- und Zeitreihenanalyse.<sup>296</sup> Im Rahmen der Untersuchungen der Algorithmen-Aversion werden häufig solche Predictive Analytics untersucht. Der in der Forecastingaufgabe dieser Untersuchung beschriebene Algorithmus basiert auf einer Regressionsanalyse (siehe Abschnitt 3.2.3).

Beim Prescriptive Analytics geht es darum, Methoden und Maßnahmen zu finden, mit denen ein Ziel optimalerweise erreicht werden – also ein Optimierungsproblem gelöst werden kann. Um diese zu statuieren, werden die Ergebnisse aus deskriptiver und prädiktiver Analyse genutzt. Dafür werden Algorithmen zur Optimierung und Simulation eingesetzt.<sup>297</sup> Seiter (2019) definiert dazu: „Optimierung und Simulation dienen der modellbasierten Identifikation jener Ausprägungen von relevanten Attributen, die zur optimalen Ausprägung der zu optimierenden Attribute führen“<sup>298</sup>.

Dieser Abschnitt kann zusammengefasst werden mit der Erkenntnis, dass Algorithmen in allen Bereichen des Forecastings verwendet werden können. Die Bedeutung von Algorithmen im Controlling zeigt sich ferner in der Tragweite der organisatorischen Gestaltung des Controllings wie auch der Anpassung von Kompetenzen von

---

<sup>293</sup> Vgl. Gronau et al. 2016, S. 472.

<sup>294</sup> Vgl. Appelbaum et al. 2017, S. 29.

<sup>295</sup> Vgl. Appelbaum et al. 2017, S. 29; Seiter 2019, S. 31.

<sup>296</sup> Vgl. Seiter 2019, S. 31.

<sup>297</sup> Vgl. Seiter 2019, S. 31.

<sup>298</sup> Seiter 2019, S. 153.

Controllern und des Rollenbildes. Algorithmen werden demnach dauerhaft eingesetzt. Dies ist elementar für den Brückenschlag zur Algorithmen-Aversion. Die vielfach erwähnte Algorithmen-Appreciation ist damit für das Forecasting von untergeordneter Relevanz, denn die wiederholte Zusammenarbeit mit Algorithmen kann nicht verhindert werden.<sup>299</sup> Im Gegenteil, organisatorische und personelle Ressourcen werden zunehmend auf die intensive und wiederholte Arbeit mit Algorithmen ausgelegt.

### 2.2.2 Einflüsse auf das Entscheidungsverhalten von Controllern

Aus dem vorangegangenen Abschnitt wird deutlich, dass das Ziel des Controllings, die Bereitstellung von finanziellen und nicht finanziellen Informationen an das Management, einem starken Wandel unterworfen ist.<sup>300</sup> Damit die Controller diesem Wandel gerecht werden können, stehen die Kompetenzprofile und Rollenbilder unter Veränderungsdruck.<sup>301</sup> Controller entwickeln sich dabei zunehmend zu Businesspartnern des Managements.<sup>302</sup> Das bedeutet, dass sie das Management mit immer weitreichenderen und wertvolleren Daten und Informationen versorgen und für die optimale Auswertung und Interpretation der Ergebnisse maßgeblich verantwortlich sind.<sup>303</sup>

Neue Kompetenzen werden in den Bereichen der Datenkompetenz, der Analysekompetenz sowie dem Technologie- und Geschäftsverständnis gesehen.<sup>304</sup> Schläfke et al. (2013) stellt dabei fest: „However, sometimes these tools – and the skills needed to execute them – lie outside the domain of traditional PMSs“<sup>305</sup>. Vor diesem Hintergrund ist auch eine Erweiterung des Kompetenzprofils von Controllern zum Data Scientist denkbar.<sup>306</sup> Es herrscht noch kein einheitliches Bild zum Grad der Weiterentwicklung des Controllers zum Data Scientist. Abbildung 17 zeigt die Überschneidung der bedeutsamsten Aufgaben von Controllern und Data Scientists.

---

<sup>299</sup> Vgl. Logg et al. 2019, S. 90; Prah und van Swol 2017, S. 691.

<sup>300</sup> Vgl. Appelbaum et al. 2017, S. 30.

<sup>301</sup> Vgl. Tiron-Tudor und Deliu 2021, S. 376.

<sup>302</sup> Vgl. Schäffer und Weber 2016, S. 13.

<sup>303</sup> Vgl. Appelbaum et al. 2017, S. 29.

<sup>304</sup> Vgl. Chamoni und Gluchowski 2017, S. 8.

<sup>305</sup> Schläfke et al. 2013, S. 111.

<sup>306</sup> Vgl. Matanovic et al. 2019, S. 54.

## Begriffsbestimmungen und Stand der Forschung

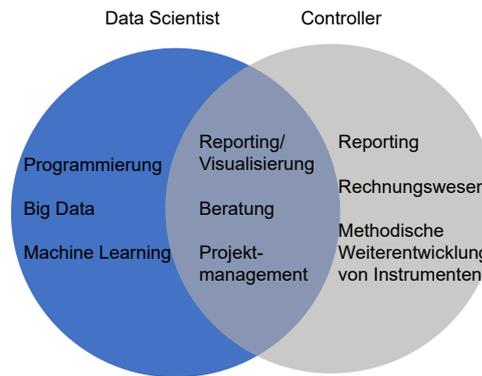


Abbildung 17: Aufgaben eines Data Scientists und eines Controllers<sup>307</sup>

Eine vollständige Überführung des Controllers zu einem Data Scientist wurde zwar in der Vergangenheit gefordert<sup>308</sup>, ist wegen der Tragweite beider Berufsgruppen nicht unbedingt zukunftsfähig.<sup>309</sup> Vielmehr sollte eine effektive Zusammenarbeit zwischen Spezialisten auf beiden Seiten etabliert werden.

Eine Diskussion um die Kompetenzentwicklung von Controllern zu Businesspartnern und den Aufbau neuer Berufsbilder, wie dem des Data Scientists, zeigen die Tragweite der Dynamik, die mit der Einführung von Algorithmen einhergeht. Schlussendlich ist es eine Weiterentwicklung und Bereitschaft zur Adaption, die auf einer individuellen Ebene gesehen werden kann. Chenhall (2003) weist vor diesem Hintergrund auf die Vergänglichkeit in der Forschung und Konzeption von Managementsteuersystemen<sup>310</sup> hin und sieht eine große Notwendigkeit, die bestehenden Ansätze, die das Verhalten von Controllern erklären können, kontinuierlich zu hinterfragen und weiterzuentwickeln.<sup>311</sup> Luft (2016) sieht Controller daher als „[...] social beings with complex and somewhat changeable motivations, not as isolated operators of stable (probably profit-maximizing) decision models“<sup>312</sup>. Infolgedessen unterliegen die Entscheidungsprozesse und der geleistete Arbeitseinsatz einer ständigen Entwicklung, die von technologischen Veränderungen wie der Einführung von Algorithmen getrieben wird.

Zu verstehen, wie das Arbeitsumfeld jeweils ausgestaltet ist, ist ein Faktor der die Algorithmen-Aversion wesentlich beeinflussen kann: „[T]he most effective incentivization program is likely to vary by environment. For this reason, it seems likely that

<sup>307</sup> Eigene Darstellung in Anlehnung an Matanovic et al. 2019, S. 61.

<sup>308</sup> Siehe dazu Heupel und Lange, S. 201–202.

<sup>309</sup> Vgl. Weißberger 2021, S. 15.

<sup>310</sup> Vgl. Chenhall 2003, S. 127.

<sup>311</sup> Vgl. Malmi und Brown 2008, S. 289; Nixon und Burns 2005, S. 260; Strauß und Zecher 2013, S. 239.

<sup>312</sup> Luft 2016, S. 9.

the implementation of successful algorithmic decision necessitates context-specific behavioral design“<sup>313</sup>. Hierzu ergänzt Kerstholt (1994): „Decision-making behaviour is considerably affected by dynamic aspects of the task environment“<sup>314</sup>. Die Einflüsse des Arbeitsumfelds scheinen demnach eine maßgebliche Rolle in der Algorithmen-Aversion spielen zu können, wurden jedoch bislang nicht untersucht. Um den Einfluss des Arbeitsumfelds untersuchen zu können, bedarf es eines etablierten theoretischen Rahmens, der das Verhalten von Forecastern bei der Einführung von Algorithmen erklären kann.

### 2.2.2.1 Incentive-Effort-Performance-Beziehung nach Bonner und Sprinkle (2002)

Bonner und Sprinkle (2002) haben einen Grundstein der experimentellen Forschung zur Arbeitsleistung im Controlling gelegt. Sie propagieren ein Framework, welches einen Rahmen für die Forschung in Settings mit *Incentives*, *Effort* und *Task Performance* aufspannt (siehe Abbildung 18) – also dem Experiment, das auch in der Literatur zur Algorithmen-Aversion angewendet wird.<sup>315</sup> Wobei die Algorithmen-Aversion im Effort hinterlegt ist und neben der Leistung zur Erstellung des eigenen Forecasts auch den Vergleich der eigenen Leistung mit der des Algorithmus beinhaltet. Finalisiert wird der Effort durch die Entscheidung im incentivierten Forecast.

Um erfassen zu können, welche Variablen in der Incentive-Effort-Performance-Beziehung wirken, werden von Bonner und Sprinkle (2002) drei weit gefasste Variablen aus 85 wissenschaftlichen Publikationen extrahiert und definiert. Die drei übergeordneten Variablen sind: *Person Variables*, *Task Variables* und *Environmental Variables* – zuletzt genannte entsprechen der Sicht auf das zu betrachtende Arbeitsumfeld. Bedeutsam zu erkennen ist, dass Bonner und Sprinkle (2002) damit zwei Aspekte adressieren, die die Herausforderungen zur Algorithmen-Aversion in dieser Arbeit maßgeblich ergänzen und eine effektive Übertragung auf das Controlling ermöglichen.

Einerseits geben sie mit dem Framework Variablen vor, die in Experimenten im Controlling von hoher Relevanz sind. Andererseits sind die angeführten Variablen nachgewiesen wirksam, wenn eine Optimierung der Task Performance angestrebt wird. Die Arbeit von Bonner und Sprinkle (2002) eröffnet damit auf Basis vielfach empirisch nachgewiesener Wirkbeziehungen den Rahmen, um die Algorithmen-Aversion

---

<sup>313</sup> Burton et al. 2019, S. 6.

<sup>314</sup> Kerstholt 1994, S. 89.

<sup>315</sup> Vgl. Bonner und Sprinkle 2002, S. 312.

## Begriffsbestimmungen und Stand der Forschung

im Arbeitsumfeld des Controllings zu untersuchen. Zusammengefasst wird die Bedeutung des Frameworks mit: „Such a model allows for full, yet parsimonious, consideration of the factors that may affect performance“<sup>316</sup>.

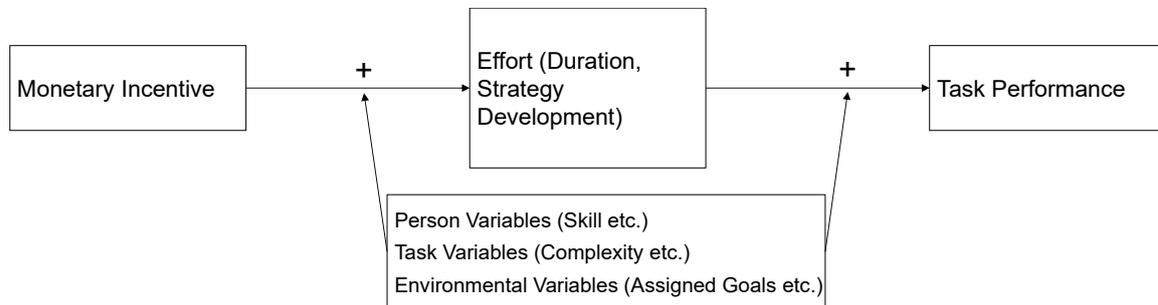


Abbildung 18: Incentive-Effort-Performance-Beziehung<sup>317</sup>

### 2.2.2.2 Eingrenzung des Untersuchungsrahmens auf das Arbeitsumfeld von Controllern

Im Folgenden werden die drei von Bonner und Sprinkle (2002) aufgestellten Variablen vorgestellt und es wird aufgezeigt, inwiefern sie bereits im Kontext der Algorithmen-Aversion adressiert wurden. Darüber hinaus wird im Detail beschrieben, welche der Environmental Variables von besonderer Relevanz für diese Untersuchung sind.

Task Variables charakterisieren die zu Grunde liegende Aufgabe. Sie definieren damit die Aufgabencharakteristik. Eine Aufgabe ist weitläufig definiert als „[...] a piece of work assigned to or demanded of a person“<sup>318</sup>. Task Variables betreffen dabei unter anderem das Framing von Aufgaben. Dies kann negatives oder positives Framing sein, wie beispielsweise die Vorhersage zur Wahrscheinlichkeit aus unternehmensbezogenen Daten, ob ein Unternehmen in den Konkurs muss (negativ), oder ob ein Unternehmen am Markt bleibt (positiv). Darüber hinaus können Task Variables beispielsweise die Komplexität einer Aufgabe betreffen. Die Komplexität erhöht sich mit dem Anspruch an die Entwicklung von Strategien zur Problemlösung oder mit steigender Unsicherheit durch die Bereitstellung von mehr oder weniger Informationen zur Problemlösung. Die Komplexität der Aufgabe in der Algorithmen-Aversion ist demnach in der Regel hoch, da neben der Analyse der Daten für den eigenen Forecast auch die Leistung des Algorithmus beurteilt werden muss. Die Aufgabenstellung wurde dabei bereits einige Male in der Forschung variiert. Niszczoła und Kaszás (2020) betonen

<sup>316</sup> Bonner und Sprinkle 2002, S. 311.

<sup>317</sup> Eigene Darstellung in Anlehnung an Bonner und Sprinkle 2002, S. 304.

<sup>318</sup> Bonner und Sprinkle 2002, S. 311.

moralische Aspekte in einer Aufgabenstellung als Hemmnis für den Einsatz von Algorithmen<sup>319</sup>, und Castelo et al. (2019) zeigen, dass subjektiv wahrgenommene Aufgaben zu einer höheren Algorithmen-Aversion führen als objektiv wahrgenommene Aufgaben.<sup>320</sup> Algorithmen sollten jedoch in allen Aufgaben eingeführt werden, in denen sie die die Forecastinggenauigkeit verbessern. Das Framing von Subjektivität zu Objektivität oder moralischer Aspekte einer Aufgabe kann in etablierten Aufgaben nicht durchgeführt werden: Einerseits wird die Aufgabe womöglich schon länger vom jeweiligen Controller durchgeführt. Andererseits werden die Aufgaben durch das Framing verzerrt oder Informationen entfernt. Die Task Variables erscheinen daher als Variablen, die in den Untersuchungen zur Algorithmen-Aversion i. e. S. konstant gehalten werden sollten.

Person Variables sind solche Variablen, die sich auf das betrachtete Individuum bzw. den Probanden beziehen. Sie sind bereits vor dem Ausführen einer Aufgabe bekannt und können im Vorfeld abgefragt und definiert werden.<sup>321</sup> Sie sind damit persönliche Charakteristika wie Motivation, Persönlichkeit, Fähigkeiten, intrinsische Motivation, Risikopräferenz oder Confidence.<sup>322</sup> Diese Variablen hängen demnach primär von den untersuchten Personen ab. Als bedeutsamste Person Variable im Kontext einer incentivierten Aufgabe wird die *Fähigkeit* einer Person gesehen.<sup>323</sup> Dies hat drei Gründe: Es wird angenommen, dass einige der anderen Person Variables unter der Fähigkeit einer Person für die Erfüllung einer bestimmten Aufgabe subsumiert werden können.<sup>324</sup> Weiter stellen die Fähigkeiten einer Person im Controlling eine entscheidende Rolle für den Erfolg von controllingbezogenen Aufgaben dar.<sup>325</sup> Außerdem spielen die Fähigkeiten eine entscheidende Rolle bei der Verbesserung der Performance. Die monetären Incentives im Incentive-Effort-Performance-Schema sollen zur bestmöglichen Leistung führen, sind jedoch abhängig von der Fähigkeit der Personen und können daher teilweise nur begrenzt auf die Leistung wirken.<sup>326</sup>

---

<sup>319</sup> Vgl. Niszczoła und Kaszás 2020, o. S.

<sup>320</sup> Vgl. Castelo et al. 2019, S. 809.

<sup>321</sup> Vgl. Bonner und Sprinkle 2002, S. 311.

<sup>322</sup> Vgl. Bonner und Sprinkle 2002, S. 312–313.

<sup>323</sup> Fähigkeit bezieht sich hier auf den angelsächsischen Ausdruck *skill*.

<sup>324</sup> Vgl. Bonner und Sprinkle 2002, S. 313.

<sup>325</sup> Vgl. Bonner und Lewis 1990, S. 1; Libby und Luft 1993, S. 425.

<sup>326</sup> Vgl. Bonner 1999, S. 386.

## Begriffsbestimmungen und Stand der Forschung

Person Variables werden bislang in der Forschung zur Algorithmen-Aversion insofern untersucht, dass gezeigt wurde, dass die Confidence ein Treiber der Algorithmen-Aversion ist. Gegenüber Laien steigt die Aversion demnach bei Experten an. Burton et al. (2019) sehen die Möglichkeit einer Reduktion der Algorithmen-Aversion durch Schulungen und ein besseres Verständnis für Algorithmen.<sup>327</sup> Dies kann einerseits durch die kognitiven Fähigkeiten von Controllern an Grenzen stoßen und ist andererseits ein individuelles Vorgehen je nach Algorithmus.

Environmental Variables, also Faktoren aus dem Arbeitsumfeld umfassen „[...] all the conditions, circumstances, and influences surrounding a person who is doing a specific task“<sup>328</sup>. Environmental Variables sind damit unabhängig von einer bestimmten Aufgabe und bestimmten Personen – können jedoch in jeder Aufgabe auftreten und auf die ausführenden Personen wirken. Die Environmental Variables umspannen ein Feld, das sich sowohl im Controlling als auch Finance, Management, in der Psychologie oder der Ökonomie wiederfindet.<sup>329</sup> Dieses zunächst recht weitläufig erscheinende Feld der Environmental Variables wird durch folgende, erprobte Variablen beschrieben:<sup>330</sup> *time pressure, accountability requirements and assigned goals, the assignment of decision rights* und *feedback*.<sup>331</sup> Bonner und Sprinkle (2002) begründen diese Auswahl der Variablen mit: „We primarily devote our attention to environmental variables that are important in accounting settings and that have been studied in combination with monetary incentives“<sup>332</sup>.

- Accountability requirements werden im Rahmen dieser Untersuchung nicht weiter betrachtet, da sie nicht mit der zugrunde liegenden Aufgabe interagieren können, ohne den Aufbau der Aufgabe wesentlich zu verändern.
- Time pressure ist eine Variable, die bislang keine Berücksichtigung in der Forschung zur Algorithmen-Aversion erfahren hat. Das heißt, die Probanden in Experimenten erhalten keine zeitliche Begrenzung. Da Zeitdruck jedoch in jeder Aufgabe auftreten kann, wird der Einfluss von Zeitdruck auf die Algorithmen-

---

<sup>327</sup> Vgl. Burton et al. 2019, S. 4.

<sup>328</sup> Bonner und Sprinkle 2002, S. 311.

<sup>329</sup> Vgl. Bonner und Sprinkle 2002, S. 311.

<sup>330</sup> Vgl. Bonner und Sprinkle 2002, S. 325–331.

<sup>331</sup> Monetäre Incentives selbst sind ebenfalls eine Environmental Variable. Sie werden jedoch in der Regel als Konstante betrachtet, können jedoch mit den weiteren Environmental Variables interagieren. Vgl. Bonner und Sprinkle 2002, S. 312.

<sup>332</sup> Bonner und Sprinkle 2002, S. 325.

Aversion in Studie 1 näher untersucht. Die detaillierte Beschreibung, welche Wirkung von Zeitdruck auf die Algorithmen-Aversion zu erwarten ist, findet sich in Abschnitt 4.2.2.

- Assigning goals, also die Zuweisung von Zielen, erfolgt im Kontext der Algorithmen-Aversion sowohl verbal als auch durch das Incentive. Das heißt, es wird ein schwieriges und spezifisches Ziel induziert. Dem gegenüber stehen in der Praxis und Theorie Ziele, die nicht direkt an eine Leistung geknüpft sind. Dabei ist bislang nicht untersucht, wie ein Ziel, welches als „do your best“, also „gib dein Bestes“ formuliert ist, in der Algorithmen-Aversion wirkt. Ein solches „do your best“-Ziel wird daher ebenfalls in Studie 1 untersucht. Die potenzielle Wirkung eines „do your best“-Ziels auf die Algorithmen-Aversion wird in 4.2.3 beschrieben.
- The assignment of decision rights liegt in den Aufgaben der Algorithmen-Aversion zugrunde, wenn die Wahl zwischen dem eigenen Forecast und dem des Algorithmus getroffen werden kann. Als wirksame Erweiterung dieser Entscheidungsrechte zur Reduktion der Algorithmen-Aversion hat sich nach Dietvorst et al. (2018) das zusätzliche Entscheidungsrecht als Eingriff in das Ergebnis des Algorithmus gezeigt.<sup>333</sup> Es sind demnach outputorientierte Entscheidungsrechte bereits untersucht worden. Diese wiederum verschlechtern den Forecast in aller Regel. Es gibt darüber hinaus bislang keine Untersuchung zu den Inputfaktoren, also inwiefern eine Algorithmen-Aversion sich verändert, wenn ein Forecaster die Entscheidungsrechte über die zu verarbeitenden Daten erhält (siehe dazu Studie 1). Welche Wirkung aus Entscheidungsrechten auf die Algorithmen-Aversion zu erwarten ist, wird in Abschnitt 4.2.4 diskutiert.
- Feedback stellt in Form von Leistungsfeedback ein wesentliches Element der Algorithmen-Aversion dar. Aus Sicht der Environmental Variables, wird dabei jedoch bislang ein zentrales Element vernachlässigt. Forecaster sind in ihrem Arbeitsumfeld stets von Kollegen und Konkurrenten umgeben. Das heißt sie ordnen ihre Leistung nicht nur gegenüber der eines Algorithmus ein, sondern auch gegenüber der ihrer Kollegen. Die Integration eines solchen Leistungs-

---

<sup>333</sup> Vgl. Dietvorst et al. 2018, S. 1155.

## Begriffsbestimmungen und Stand der Forschung

feedbacks in Relation zu Kollegen wird als Peer-Feedback bzw. RPI beschrieben.<sup>334</sup> Mit Hinblick auf die Steuerung der Leistung ist RPI ein wirksames Feedbackinstrument und daher von großer Relevanz.<sup>335</sup> RPI kann je nach Leistung eines Forecasters motivierend oder auch demotivierend wirken.<sup>336</sup> Außerdem stellt sie nicht nur eine zusätzliche Komponente zum bestehenden Leistungsfeedback dar, sondern kann eine Wirkung auf sozialer Ebene entfalten.<sup>337</sup> Die Einordnung von RPI und die Ableitung von Hypothesen zur Wirkung von RPI in das theoretische Modell der Algorithmen-Aversion erfordern daher eine umfassende Betrachtung bestehender empirischer Befunde. Dem wird in Studie 2 Rechnung getragen. Welche Wirkung von RPI auf die Algorithmen-Aversion erwartet wird, ist in Abschnitt 5.2 zu finden.

Untenstehend fasst Abbildung 19 die relevanten Variablen nochmals zusammen und ordnet sie den in Kapitel 4 (Studie 1) und Kapitel 5 (Studie 2) folgenden empirischen Studien zu. Studie 1 behandelt Zeitdruck, „do your best“-Ziele und die Entscheidungsrechte für Eingabeinformationen in der Interaktion mit einem Algorithmus. Studie 2 erweitert die empirischen Erkenntnisse zur Algorithmen-Aversion um den Einfluss von RPI.

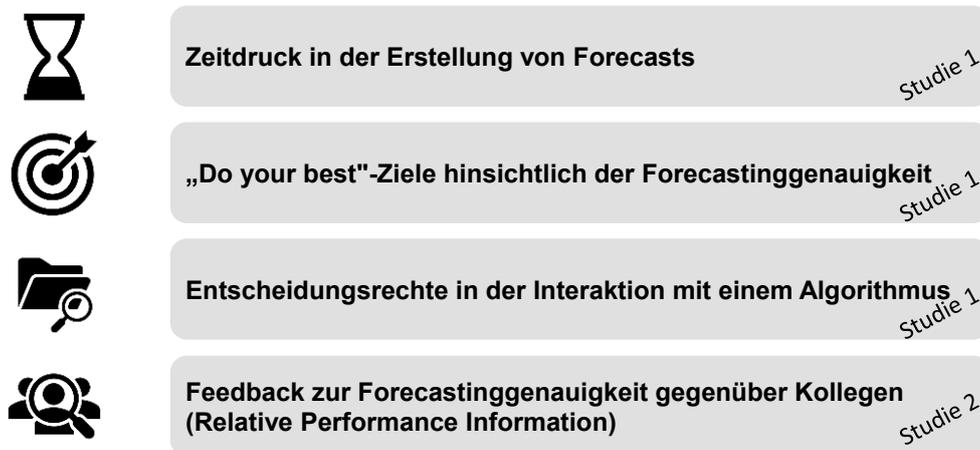


Abbildung 19: Zuordnung der relevanten Variablen aus dem Arbeitsumfeld zu den empirischen Studien dieser Untersuchung<sup>338</sup>

<sup>334</sup> Vgl. Hannan et al. 2013, S. 553.

<sup>335</sup> Vgl. Newman et al. 2021; Tafkov 2013, S. 327.

<sup>336</sup> Vgl. Mahlendorf et al. 2014, S. 331.

<sup>337</sup> Vgl. Alder 2007, S. 157; Festinger 1954, S. 117; Kluger und DeNisi 1996, S. 254.

<sup>338</sup> Eigene Darstellung.

### 3 Grundlagen zur empirischen Erhebung

Die empirischen Erhebungen in dieser Arbeit folgen der bestehenden Forschung zur Algorithmen-Aversion. Die Diskussion im Stand der Forschung hat gezeigt, dass die Algorithmen-Aversion im Arbeitsumfeld von Forecastern nur unverzerrt untersucht werden kann, wenn andere potenzielle Einflussfaktoren ausgeschlossen werden können. Damit bei der Einführung von Faktoren aus dem Arbeitsumfeld keine ungewollten Verzerrungen entstehen, wird das Experimentalsetting von Dietvorst et al. (2015) repliziert und erweitert. Repliziert wird dabei Studie 3b, da diese die höchste statistische Power und damit die am weitreichendsten gesicherten Erkenntnisse aufweist. Es wird daher, wie in Studie 3b, ein Online-Experiment durchgeführt.<sup>339</sup> Das Experiment wird auf der in Deutschland ansässigen Plattform *SoSci Survey* programmiert.<sup>340</sup> Die Untersuchung wird Dietvorst et al. (2015) folgend auf der Plattform *MTurk* durchgeführt (siehe Abschnitt 3.1.2). So kann sichergestellt werden, dass keine Verzerrungen aus divergierenden persönlichen Merkmalen entsteht. Die Forecastingaufgabe, der Algorithmus, die Incentivierung und die abgefragten Variablen folgen ebenfalls im Wesentlichen Studie 3b. Der Ablauf des Experiments und etwaige notwendige Änderungen im Ablauf werden in Abschnitt 3.2 beschrieben. Die Grundlagen zur Datenauswertung finden sich in Abschnitt 3.3. Abbildung 20 fasst den Aufbau von Kapitel 3 zusammen.

---

<sup>339</sup> Vgl. Dietvorst et al. 2015, S. 119.

<sup>340</sup> Zu erreichen unter: <https://www.sosicurvey.de/> (zuletzt geprüft am 31.05.2022).

# Grundlagen zur empirischen Erhebung

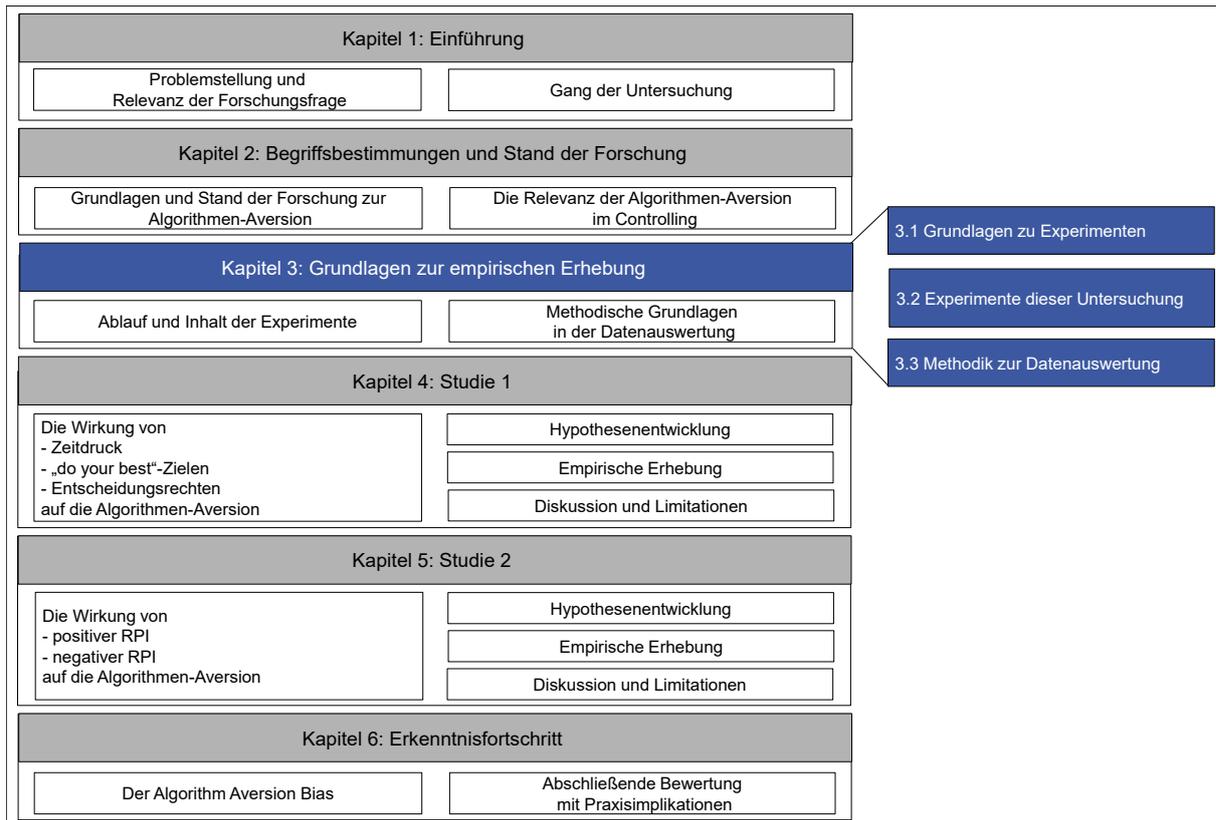


Abbildung 20: Aufbau von Kapitel 3<sup>341</sup>

## 3.1 Grundlagen zu Experimenten

### 3.1.1 Experimente und Replikation

Die Hypothesen in dieser Untersuchung folgen einer Ursache-Wirkungs-Beziehung.<sup>342</sup> Das heißt, sie suggerieren einen gerichteten kausalen Einfluss auf den Einsatz von Algorithmen in Forecasts. Um solche Kausalhypothesen zu überprüfen, stehen explanative Studientypen wie Experimente zur Verfügung.<sup>343</sup> Experimente sind ein im wissenschaftlichen Erkenntnisgewinn zur Überprüfung von Kausalhypothesen weit verbreitetes Instrument der Datengewinnung. Sarris (1990) versteht unter einem Experiment „[...] einen systematischen Beobachtungsvorgang, aufgrund dessen der Untersucher das jeweils interessierende Phänomen erzeugt sowie variiert und dabei gleichzeitig systematische und/oder unsystematische Störfaktoren durch hierfür geeignete

<sup>341</sup> Eigene Darstellung.

<sup>342</sup> Vgl. Eifler 2014, S. 198.

<sup>343</sup> Vgl. Döring und Bortz 2016, S. 193.

Techniken ausschaltet bzw. kontrolliert“<sup>344</sup>. Die Merkmale von Experimenten werden im Folgenden beschrieben und in den Kontext dieser Untersuchung eingeordnet.<sup>345</sup>

Grundsätzlich kann in Labor- und Feldexperimente unterschieden werden.<sup>346</sup> Laborexperimente tragen ihren Namen wegen der laborähnlichen Bedingungen, die für die Durchführung des Experiments geschaffen werden. Dies sind häufig eigens für das Experiment ausgestattet und hergerichtete Räumlichkeiten. Feldexperimente hingegen werden in der natürlichen Umgebung der Probanden durchgeführt. Da die Experimente in dieser Studie auf MTurk<sup>347</sup> durchgeführt werden und die MTurker dort einer sporadischen Arbeit nachgehen, kann das Experiment als Feldexperiment klassifiziert werden.

Kausalschlüsse in einem Experiment werden durch den Vergleich von Experimentalgruppen gezogen. Zu diesem Zweck müssen zwei oder mehr Gruppen gebildet werden. Diese Gruppen sollen möglichst uneingeschränkt vergleichbar sein.<sup>348</sup> Hierzu müssen die Gruppen randomisiert sein.<sup>349</sup> Das heißt, die Teilnehmer werden den Gruppen zufällig zugewiesen. Dadurch kann auf systematische, personenbezogene oder untersuchungsbedingte Störgrößen hin kontrolliert werden. Die Randomisierung soll sicherstellen, dass die Probanden zufällig auf die Kontroll- und Treatmentgruppen verteilt werden. Die Gruppen werden von der aktiven Manipulation durch den Forscher erzeugt.<sup>350</sup> Die manipulierte Gruppe wird als Treatmentgruppe bezeichnet. Die Manipulation meint eine systematische Ungleichbehandlung der Gruppen durch Veränderung einer oder mehrerer unabhängiger Variablen.<sup>351</sup> Gemessen wird im Anschluss der Effekt, den die Manipulation auf eine oder mehrere abhängige Variablen hat.<sup>352</sup> Dieser Effekt ist demnach der Unterschied in den betrachteten Gruppen durch das Treatment oder die vom Forscher herbeigeführte Variation.<sup>353</sup> Damit der Effekt auch der gewünschten Ursache-Wirkungs-Beziehung zugerechnet werden kann, müssen

---

<sup>344</sup> Sarris 1990, S. 129.

<sup>345</sup> Vgl. Sarris 1990, S. 129–130.

<sup>346</sup> Vgl. Eifler 2014, S. 205–207.

<sup>347</sup> Siehe Abschnitt 3.1.2.

<sup>348</sup> Vgl. Döring und Bortz 2016, S. 193.

<sup>349</sup> Vgl. Döring und Bortz 2016, S. 193; Ist keine Randomisierung möglich, wird von einem *Quasi-Experiment* gesprochen. Siehe hierzu Döring und Bortz 2016, S. 193.

<sup>350</sup> Vgl. Döring und Bortz 2016, S. 193.

<sup>351</sup> Werden mehrere unabhängige Variablen manipuliert wird von einem *faktoriellen Design* gesprochen. Siehe dazu Baur und Blasius 2014, S. 204.

<sup>352</sup> Vgl. Döring und Bortz 2016, S. 193.

<sup>353</sup> Vgl. Eifler 2014, S. 198.

alternative Einflussfaktoren auf die Probanden möglichst ausgeschlossen werden. Aufgrund dieser kontrollierten Randomisierung und Manipulation gelten Experimente nach Döring und Bortz (2016) als *Königsweg der Erkenntnis* zur Überprüfung von Kausalhypothesen.<sup>354</sup>

Um die bestmögliche Vergleichbarkeit der Kontroll- und Treatmentbedingungen sicherzustellen, werden Kontrolltechniken angewendet. Hierbei wird unterschieden zwischen Kontrolltechniken, die sich auf den experimentellen Ablauf beziehen, und solchen, die sich auf die Probanden beziehen.<sup>355</sup> Hinsichtlich des Ablaufs ist auf eine hohe Konstanzhaltung zwischen den einzelnen Gruppen zu achten. Für den Ablauf des Experiments in dieser Untersuchung wird zunächst das Experiment der Kontrollgruppe gemäß Dietvorst et al. (2015) nachgebildet.<sup>356</sup> Die Kontrollgruppen in dieser Studie sind demnach Replikationen. Eine Replikation ist definiert als „[...] a duplication of a previously published empirical study that is concerned with assessing whether similar findings can be obtained upon repeating the study“<sup>357</sup>. Durch diese Replikation wird die interne Validität des Experiments gesichert.<sup>358</sup> Replikationsstudien erhöhen die interne Validität in zweifacher Weise. Zum einen lassen sie einen Vergleich der Ergebnisse mit Erkenntnissen zur Algorithmen-Aversion zu und geben einen Anhaltspunkt, ob die Ausgangsdiskussion der robusten Algorithmen-Aversion noch Bestand hat, wie es zum Zeitpunkt der bisherigen Erhebungen im Jahr 2015 der Fall war. Zum anderen werden dadurch Kontrollbedingungen geschaffen, die zeitgleich wie die Treatmentbedingungen erhoben werden. Das heißt, es wird auf kurzfristige Einflussfaktoren, wie beispielsweise die Uhrzeit der Erhebung, und auf langfristige Faktoren, wie eine demografische Veränderung der Stichprobe, kontrolliert. Die Treatments in den Experimenten werden minimalinvasiv in den Ablauf der Kontrollgruppen integriert (siehe Abschnitt 4.3.1 und 5.3.1).

### 3.1.2 Plattform und Teilnehmer der Experimente

Um möglichst große Kontrolle über die Manipulation in den Experimentalstudien zu haben, wird die Stichprobe aus Dietvorst et al. (2015) repliziert. Die Probanden werden

---

<sup>354</sup> Vgl. Döring und Bortz 2016, S. 193.

<sup>355</sup> Vgl. Eifler 2014, S. 201.

<sup>356</sup> Da zwischen Studie 1 und Studie 2 ein Jahr vergangen ist, wird sowohl für Studie 1 als auch für Studie 2 die Kontrollbedingung erhoben.

<sup>357</sup> Hubbard und Armstrong 1994, S. 236.

<sup>358</sup> Vgl. Evanschitzky und Armstrong 2010, S. 4.

demnach mittels *MTurk* akquiriert. Die auf *MTurk* aktiven Arbeiter (sogenannte *Clickworker*) werden im Folgenden als *MTurker* bezeichnet. *MTurker* arbeiten online an Aufgaben, die über die Plattform *MTurk* zur Verfügung gestellt werden. Entsprechend ist ein Experiment mit *MTurkern* ein incentiviertes Setup, in dem das Incentive auch die tatsächliche Vergütung in Form eines Arbeitslohns darstellt. Mehrere Studien zeigen, dass Ergebnisse von Präsenz- und Onlineteilnehmern von Studien, Experimenten und Umfragen ähnlich ausfallen und auf qualitativer Ebene vergleichbar sind.<sup>359</sup> Dadurch werden Online-Tools, wie *MTurk*, zunehmend auch für wissenschaftliche Umfragen und Experimente genutzt. Eine Gegenüberstellung der Vor- und Nachteile dieser Methodik wird im Folgenden umrissen.

Casler et al. (2013) stellen vor allem heraus, dass Online-Panels eine größere demografische Vielfalt aufweisen als übliche Testgruppen, die sich in häufig aus College-Studenten zusammensetzen.<sup>360</sup> Im Internet erreichen die Studien breitere Bevölkerungsgruppen, die verschiedene Altersgruppen sowie ethnische und ökonomische Verhältnisse abbilden. Es können kontrolliert mehrere Nationalitäten miteinbezogen werden. Ergebnisse aus Online-Studien gewinnen somit gegenüber Präsenzbefragungen an Repräsentativität. Außerdem erhöht die größere Teilnehmerzahl die Gültigkeit der Ergebnisse.<sup>361</sup> Weiterhin profitieren Wissenschaftler vor allem von Kosten- und Zeitersparnissen. Eine große Zahl an Rückläufen kann durch *MTurk* sogar binnen weniger Stunden generiert werden und auch der Zeit- und Kostenaufwand für die vorgelegte Akquise von Testpersonen entfällt. Monetäre Ersparnisse ergeben sich zudem dadurch, dass beispielsweise keine Raum- oder Reisekosten entstehen. Casler et al. (2013) stellen weiter heraus, dass Online-Experimente und Fragebögen den Vorteil besitzen, immer exakt gleiche Voraussetzungen für die Teilnehmer zu schaffen.<sup>362</sup> Durch die auf Internetplattformen geschaffene Anonymität gehen sie davon aus, dass Teilnehmer die Fragen ehrlicher beantworten als bei Experimenten mit physisch anwesenden Teilnehmern. Auch Motivationsdefizite treten auf *MTurk* durch die vorhandene Bezahlung nicht mehr oder weniger häufig auf als bei traditionellen Forschungsteilnehmern.

---

<sup>359</sup> Vgl. Bartneck et al. 2015; Casler et al. 2013; Goodman et al. 2013.

<sup>360</sup> Vgl. Casler et al. 2013, S. 2156.

<sup>361</sup> Siehe hierzu auch Smyth und Pearson 2011, S. 11.

<sup>362</sup> Vgl. Casler et al. 2013, S. 2157.

Smyth und Pearson (2011) verweisen auf die digitalen Vorteile.<sup>363</sup> Bei der Datenerhebung im Internet kann nicht nur mit zeitigen Rückmeldungen gerechnet werden, sondern gibt es sogar die Möglichkeit der Erhebung in Echtzeit. Auch Zufallsgeneratoren lassen sich online einfacher und zuverlässiger abbilden. Im nächsten Schritt sind die Ergebnisse dann bereits in digitaler Form vorhanden und können ohne, dass eine erneute Eingabe erforderlich wäre, weiterverarbeitet werden. Somit lassen sich schließlich menschliche Fehler reduzieren. Viele Online-Tools erlauben den Forschern eine komplexere Ausarbeitung ihrer Aufgabenstellungen sowie mehr Möglichkeiten zur visuellen Gestaltung. Sollten während der Datenerhebung Fehler auftreten, können die Aufgaben leichter erneut verändert und bearbeitet werden als im Labor. Werden weiterführende Forschungen angestrebt, die auf das Experiment aufbauen oder es replizieren wollen, können Codes der Experimente einfach weitergegeben werden und müssen nicht erneut aufwändig nachgebaut werden.

Online-Tools können auch für Teilnehmer der Experimente günstigere Bedingungen schaffen. Dadurch, dass Forscher und Teilnehmer unbekannt bleiben, wird der Druck auf den Teilnehmer verringert und der Verlauf des Experiments kann nicht vom Forscher gewollt oder ungewollt beeinflusst werden.<sup>364</sup> Durch die Einstellung des Experiments auf MTurk, könnte – die ethischen Bedenken außer Acht gelassen – den Probanden sogar gänzlich verschwiegen werden, dass es sich überhaupt um ein wissenschaftliches Experiment handelt. Somit kann Unvoreingenommenheit erreicht werden.<sup>365</sup> Die Online-Bearbeitung erlaubt den Teilnehmern außerdem, ihr eigenes Tempo zu wählen.<sup>366</sup>

Zusätzlich zu den bereits genannten Vorteilen von MTurk und weiteren Online-Mitteln zur Durchführung von Experimenten, merken Mason und Suri (2012) an, dass durch deren Anwendung viel mehr Wissenschaftler die Möglichkeit erhalten, Fragestellungen empirisch testen zu lassen.<sup>367</sup> Vor allem an kleineren Universitäten oder Forschungseinrichtungen gibt es dafür oft weder die finanziellen und materiellen Mittel noch den nötigen Zugang zu Teilnehmern. Für MTurk spricht zudem die einfache und sichere Zahlungsabwicklung, die durch den Anbieter Amazon geregelt ist und somit

---

<sup>363</sup> Vgl. Smyth und Pearson 2011, S. 11.

<sup>364</sup> Vgl. Crump et al. 2013, S. 2.

<sup>365</sup> Vgl. Horton et al. 2011, S. 400.

<sup>366</sup> Vgl. Smyth und Pearson 2011, S. 15.

<sup>367</sup> Vgl. Mason und Suri 2012, S. 3.

einerseits dem Forscher die Entscheidung über geeignete Methoden abnimmt und andererseits dem Teilnehmer die nötige Sicherheit gibt.

Die Nachteile der Anonymität des Internets ergeben sich auch bei Online-Umfragen und Experimenten sowie beim Anbieter MTurk. Die Identität der Teilnehmer und deren angegebene demografische Daten können letztlich nicht vollends überprüft werden.<sup>368</sup> Ebenso besteht die Möglichkeit, dass mehrere Teilnehmer denselben Account nutzen oder, dass eine Person mehrere Accounts nutzt. Im Extremfall kann es sogar sein, dass das Experiment gar nicht von einem Menschen bearbeitet wird, sollten Bots hinter den Accounts stecken. Es ist daher notwendig, entsprechende Filterfragen, Aufmerksamkeitschecks und logische Überprüfungen der Antworten in den Aufbau des Experiments und die anschließende Datenaufbereitung zu integrieren. Um einigen dieser Herausforderungen entgegenzuwirken, wird in der Literatur geraten, Testfragen einzubauen. Horton et al. (2011) schlagen vor, Fragen einzubauen, durch die festgestellt werden kann, ob alle Teilnehmer die Fragestellung gleich und inhaltlich korrekt auffassen.<sup>369</sup> Dies kann vor allem deswegen nötig werden, da es den Probanden nicht möglich ist, Rück- und Verständnisfragen an den Forscher zu stellen. Auch Testfragen, die eine Bearbeitung durch Bots ausschließen, sind sinnvoll und können leicht eingebaut werden.<sup>370</sup> Schließlich empfehlen sie, die Aufmerksamkeit der Teilnehmer durch den Einbezug entsprechender Fragen zur Überprüfung der Aufmerksamkeit zu stellen.<sup>371</sup>

Bei Experimenten im Internet ist eine direkte *face-to-face-Kommunikation* nicht möglich und somit erlauben sie zumeist auch keine Rückfragen.<sup>372</sup> Da für die Forscher auch das Umfeld und die Umstände, unter denen der Teilnehmer die Aufgaben oder Fragebögen bearbeitet, nicht ersichtlich sind, können die Ergebnisse von Variablen beeinflusst werden, die dem Wissenschaftler verborgen bleiben.<sup>373</sup> Auch die bewusste Beeinflussung der Umwelt oder bestimmter Variablen im Sinne des Experiments durch den Forscher ist nicht möglich.<sup>374</sup> Generell kann nicht jedes Experiment per se auch

---

<sup>368</sup> Vgl. Berinsky et al. 2012, S. 352; Horton et al. 2011, S. 405.

<sup>369</sup> Vgl. Horton et al. 2011, S. 404.

<sup>370</sup> Vgl. Crump et al. 2013, 2.

<sup>371</sup> Vgl. Goodman et al. 2013, S. 216.

<sup>372</sup> Vgl. Horton et al. 2011, S. 405.

<sup>373</sup> Vgl. Bartneck et al. 2015, 3.

<sup>374</sup> Vgl. Horton et al. 2011, S. 404.

online durchgeführt bzw. auf das Setting im Internet übertragen werden. Dies ist beispielsweise der Fall, wenn eine Interaktion zwischen mehreren Probanden oder dem Forscher und den Probanden zwingend notwendig ist oder, wenn das augenscheinliche Verhalten des Teilnehmers beobachtet werden soll.<sup>375</sup>

Paolacci et al. (2010) äußern Bedenken darüber, ob mit MTurk die gewünschte Population repräsentiert werden kann und die Befragten die Experimente ernst nehmen.<sup>376</sup> Beide Aspekte müssen auch bei Präsenzteilnehmern bedacht werden, jedoch prägen sie sich online noch stärker aus. 57 % der MTurker sind US-Amerikaner. Nach ihnen bilden Inder die zweitgrößte Gruppe mit 32 %. Folglich repräsentieren nur 11 % der Teilnehmer Länder außerhalb der USA und Indien.<sup>377</sup> Je nach Zielgruppe des Experiments kann es demnach durchaus sein, dass die Teilnehmerdemografie der MTurker unpassend erscheint. Viele MTurker nutzen MTurk aus finanziellen Gründen und versuchen deshalb, so viele HITs wie möglich in kürzester Zeit zu erledigen. Darunter kann die sorgfältige Bearbeitung leiden und es kann sein, dass die Angaben und Fragestellungen von den Teilnehmern nicht genau gelesen werden, was wiederum zu schlechten Ergebnissen führt.<sup>378</sup>

Die negativen Effekte durch MTurk können durch den Replikationscharakter ausgeklammert werden. Potenzielle Verzerrungen die daraus resultieren, bestehen damit sowohl in den Experimenten von Dietvorst et al. (2015) als auch in den Experimenten dieser Arbeit. Für die externe Validität und die Übertragung der Erkenntnisse in die Praxis, müssen sie jedoch berücksichtigt werden.

### 3.2 Experimente dieser Untersuchung

Jede Seite des Experiments, die den Teilnehmern präsentiert wurde, ist unter Berücksichtigung der Differenzierung in Kontroll- und Treatmentgruppen in Anhang 6 zu finden. Im Folgenden wird der Ablauf des Experiments skizziert.

Das Experiment beginnt mit einem Hinweis zur universitären Herkunft des durchführenden Forschers, der voraussichtlichen Dauer der Durchführung und dem Vermerk, dass die Teilnahme freiwillig ist und ein Ausstieg zu keinem Zeitpunkt Konsequenzen für die Probanden hat. Die freiwillige Basis ist insofern ein bedeutsamer

---

<sup>375</sup> Vgl. Horton et al. 2011, S. 405.

<sup>376</sup> Vgl. Paolacci et al. 2010, S. 411.

<sup>377</sup> Vgl. Ross et al. 2010, S. 2864.

<sup>378</sup> Vgl. Paolacci et al. 2010, S. 414.

Aspekt, da die Plattform MTurk die Möglichkeiten zur Ex-post-Bewertung von Teilnehmern bereitstellt und hier keinerlei Druck durch drohende negative Bewertungen erzeugt werden soll. Zuletzt wird auf der ersten Seite die MTurk-ID abgefragt, damit eine spätere Verlinkung von Datensatz und Teilnehmer auf der Plattform hergestellt werden kann. Die Kenntnisse über die MTurk-ID wiederum lassen keinerlei Rückschlüsse auf die dahinterstehende Person zu und dienen lediglich der Freigabe des erreichten Incentives an eben diese MTurk-ID.

Die zweite Seite dient der Überprüfung der Aufmerksamkeit der Probanden. Die Überschrift auf dieser Seite suggeriert, dass die Probanden hier ihre Erfahrung mit MTurk vermerken sollen. Beim genauen Lesen des Texts wird jedoch eine andere Anweisung gegeben. Die Teilnehmer sollen anstelle ihrer Erfahrung mit MTurk das Wort *dolphin* eintippen. Falls nicht *dolphin* eingegeben wird, werden die Teilnehmer auf Seite 3 weitergeleitet und das Experiment wird beendet.

Seite 4 dient einer Filterfrage und ist eine Erweiterung des originären Experiments von Dietvorst et al. (2015). Hier wird abgefragt, ob die Teilnehmer bereits an einem Experiment mit derselben Aufgabenstellung teilgenommen haben (siehe Abbildung 21).

**In this task you will estimate the rank of 1 U.S. state in terms of the number of airline passengers who departed from that state in 2011.**

Have you ever participated in a survey with this task?

- Yes, I have participated in this task before
- No, I have not participated in this task before

Abbildung 21: Filterfrage zur Teilnahme an ähnlichen Experimenten<sup>379</sup>

Diese Frage appelliert an die Ehrlichkeit der Teilnehmer. Eine strategische falsche Antwort ist jedoch nicht zu erwarten, da die Teilnehmer den Hintergrund der Frage nicht kennen. Es ist auf MTurk Usus, dass ein Forscher mehrere Studien durchführt. Dabei kann es notwendig sein, dass Probanden Vorwissen als Qualifikation aufbauen. In dem Fall würde eine positive Antwort nicht zum Ausschluss, sondern zur Qualifikation für eine andere Studie führen. Es ist demnach der ehrliche Weg, der hier von den Teilnehmern erwartet werden kann.

---

<sup>379</sup> Eigene Darstellung (Auszug aus dem Experiment).

## Grundlagen zur empirischen Erhebung

Im weiteren Verlauf wird den Teilnehmer erklärt, welche konkrete Aufgabenstellung sie lösen müssen (siehe Abschnitt 3.2.1) und dass sie bei dieser Aufgabe leistungsabhängig vergütet werden (diese Seite wurde für die Treatmentgruppe mit dem „do your best“-Ziel ausgeklammert, siehe dazu Abschnitt 3.2.2). Darauf aufbauend wird den Teilnehmern je nach Gruppenzuteilung das Treatment vorgestellt und die Einteilung in einen der beiden Pfade mit oder ohne Leistungsfeedback wird definiert. Auf die Kommunikation des Treatments folgt eine weitere Aufmerksamkeitsprüfung, in der die Teilnehmer einen im Vorfeld definierten Satz wiedergeben müssen. Anschließend werden für die Gruppen der Algorithmen-Aversion die zehn Übungsforecasts gestartet. Nach einer Wiederholung der Beschreibung der Daten für die nun folgende Aufgabe und der Incentivierung werden die Teilnehmer dazu aufgefordert, sich für einen Forecast im incentivierten Forecast zu entscheiden (siehe Abbildung 22):

**Would you like your estimated rank or the model's estimated rank to determine your bonus?**

Use only the statistical model's estimated rank to determine my bonus.



Use only my estimated rank to determine my bonus.



Abbildung 22: Entscheidung im incentivierten Forecast<sup>380</sup>

Bevor der incentivierte Forecast durchgeführt wird, werden die Fragen zur Confidence in jeweils randomisierter Weise abgefragt (siehe Abschnitt 3.2.4). Damit es keine Verzerrungen zum experimentellen Ablauf von Dietvorst et al. (2015) gibt, werden zusätzliche Fragen gestellt, diese werden jedoch in der späteren Untersuchung nicht weiter betrachtet.<sup>381</sup> Es wurde außerdem eine Verständnisfrage als Filterfrage integriert. In dieser Frage wird die Einschätzung zur Leistung des Algorithmus thematisiert. Abschnitt 3.2.3 zeigt die Performance des Algorithmus. Der schlechteste Forecast des eingesetzten Algorithmus verfehlt den zu prognostizierenden Wert um 17 Ränge. Um pessimistische Forecaster nicht auszuschließen, wurde hier konservatives Vorgehen gewählt. Dazu wurden auf diese 17 Ränge zwei aufgerundete Standardabweichungen addiert. Teilnehmer, die daher eine Forecastinggenauigkeit des Algorithmus von schlechter als 27 Rängen angegeben haben, wurden ausgeschlossen.<sup>382</sup>

---

<sup>380</sup> Eigene Darstellung (Auszug aus dem Experiment).

<sup>381</sup> Siehe Abschnitt 3.2.4.

<sup>382</sup> Dieser Ausschluss von Teilnehmer anhand Verständnisfragen ist ein notwendiges Mittel um unaufmerksame Teilnehmer auszuschließen. Siehe dazu auch Berger et al. 2021, S. 63 oder Backhaus et al. 2021, S. 53.

Im Anschluss muss jeder Teilnehmer, unabhängig davon, wie er sich für den incentivierten Forecast entschieden hat, einen Forecast abgeben. Darauf aufbauend werden die soziodemografischen Daten abgefragt.

Abschließend wurde den Probanden für ihre Teilnahme gedankt und ein individueller Code zur Validierung der Teilnahme ausgegeben. Dieser Code auf der letzten Seite stellt sicher, dass der Teilnehmer den Fragebogen durchlaufen hat und dient als zweite Absicherung der korrekten Verlinkung von Datensatz und MTurk-ID.

### 3.2.1 Forecastingaufgabe

Die Teilnehmer am Experiment bekamen die Aufgabe, den Rang eines U.S.-Bundesstaats nach abfliegenden Passagieren zu prognostizieren (siehe Abbildung 23):

In this task you will estimate the rank of 1 U.S. state in terms of the number of airline passengers who departed from that state in 2011.

The state with the largest number of departing passengers is ranked 1st and the state with the smallest number of departing passengers is ranked 50th.

Abbildung 23: Aufgabenstellung im Experiment<sup>383</sup>

Im Anschluss wurde ihnen die Übersicht in Abbildung 24 zur Verfügung gestellt. Sie bekamen eine Liste mit fünf verschiedenen Informationen über den relevanten U.S.-Bundesstaat. Die Aufgabe bestand darin, den Rang (1 bis 50) einzelner U.S.-Bundesstaaten in Bezug auf die Anzahl der Fluggäste, die 2011 von diesem Staat abgeflogen sind, vorherzusagen. Rang 1 bedeutet, dass dieser Staat die meisten abfliegenden Fluggäste verzeichnen konnte, während Rang 50 der Staat mit den wenigsten abfliegenden Fluggästen ist.

Für ihren Forecast erhielten die Teilnehmer folgende Informationen über den jeweiligen Staat: Die Anzahl der großen Flughäfen (nach Definition des Verkehrsministeriums), die Einwohnerzahl von Rang 1 bis 50 (laut Volkszählung 2010), die Anzahl der Counties (1 bis 50), das durchschnittliche Haushaltseinkommen 2008 (1 bis 50) und die Höhe der staatlichen Ausgaben für Reisen (1 bis 50). Abbildung 24 zeigt die Darstellung der Informationen im Experiment.

---

<sup>383</sup> Eigene Darstellung (Auszug aus dem Experiment).

## Grundlagen zur empirischen Erhebung

### Number of Major Airports

The number of major airports in the state as determined by the Bureau of Transportation. All states have smaller airports that this number does not account for

### Census Population Rank - 2010

The state's rank in terms of population in 2010 from the U.S. Census Bureau (1 = most populated U.S. state; 50 = least populated U.S. state)

### Number of Counties Rank

The state's rank in terms of its number of counties (1 = U.S. state with the most number of counties; 50 = U.S. state with the least number of counties)

### Median Household Income Rank - 2008

The state's rank in terms of median household income in 2008 from the U.S. Census Bureau (1 = U.S. state with the highest median income; 50 = U.S. state with the lowest median income)

### Domestic Travel Expenditure Rank - 2009

The state's rank in terms of money spent by U.S. citizens traveling to the state in 2009 from the U.S. travel association (1 = U.S. state with the most incoming expenditures; 50 = U.S. state with the least incoming expenditures)

Abbildung 24: Informationen zur Aufgabe des Experiments<sup>384</sup>

Abbildung 25 zeigt ein Beispiel für die in der Studie genutzten Daten während der Übungsforecasts und des incentivierten Forecasts. Alle Daten wurden zufällig und ohne Zurücklegen aus den 50 U.S.-Bundesstaaten ausgewählt.

Number of Major Airports	0
Census Population Rank - 2010	31
Number of Counties Rank	16
Median Household Income Rank - 2008	50
Domestic Travel Expenditure Rank - 2009	32

Abbildung 25: Beispielhafte Darstellung der Information zu einem U.S.-Bundesstaat<sup>385</sup>

Neben diesen Informationen zur Datengrundlage wurde den Teilnehmern mitgeteilt, dass sie eine Vorhersage zum Rang des Staats von einem Algorithmus erhalten würden, der „developed by transportation analysts“ ist.<sup>386</sup> Es gab darüber hinaus die Information, dass der Algorithmus dieselben Informationen wie die Teilnehmer erhalten hat und keine weiteren Informationen in seinen Forecast einfließen lassen kann.

Das Leistungsfeedback wurde in den zehn Übungsforecasts wie in Abbildung 26 präsentiert. Es ist demnach eine Aufzählung des eigenen Forecasts, des Forecasts des Algorithmus und des wahren Werts.

<sup>384</sup> Eigene Darstellung (Auszug aus dem Experiment).

<sup>385</sup> Eigene Darstellung (Auszug aus dem Experiment).

<sup>386</sup> Siehe Anhang 6.

**Your Prediction:**

**Model's Prediction:**

**State's Actual Rank:**

Abbildung 26: Darstellung des Leistungsfeedbacks<sup>387</sup>

### 3.2.2 Incentivierung

Die Incentivierung der Teilnehmer erfolgt zweistufig. Für den Abschluss des Experiments erhielt jeder Teilnehmer 1 \$. Darüber hinaus gab es eine Bonuszahlung bzw. eine leistungsabhängige Vergütung: Für eine perfekte Vorhersage erhielten die Teilnehmer zusätzlich 1 \$. Für jede Fehlereinheit wurde dieser Betrag um 0,15 \$ reduziert. Die Fehlereinheit in jeder Vorhersage ist das Maß für die Forecastinggenauigkeit. Tabelle 2 stellt die Bezahlregel dar. Die Teilnehmer aus den Gruppen mit einem „do your best“-Ziel erhalten grundsätzlich den vollen Bonus von 1 \$, sofern sie das Experiment beenden.<sup>388</sup>

Tabelle 2: Modus für die Bonuszahlung

Forecastinggenauigkeit	Bonus in \$
Forecast stimmt exakt überein	1
Forecast weicht einen Rang ab	0,85
Forecast weicht zwei Ränge ab	0,7
Forecast weicht drei Ränge ab	0,55
Forecast weicht vier Ränge ab	0,4
Forecast weicht fünf Ränge ab	0,25
Forecast weicht sechs Ränge ab	0,1
Forecast weicht um sieben oder mehr Ränge ab	0,01

### 3.2.3 Algorithmus

Die Replikation des Experiments von Dietvorst et al. (2015) beinhaltet die Replikation des Algorithmus bzw. den Forecasts des Algorithmus in das vorliegende Experiment.

<sup>387</sup> Eigene Darstellung (Auszug aus dem Experiment).

<sup>388</sup> Die Experimente in dieser Arbeit wurden finanziell unterstützt durch die Péter Horváth-Stiftung. Hierfür sei an dieser Stelle nochmals gedankt.

Zu diesem Zweck wurden die Forecasts des Algorithmus von Berkeley J. Dietvorst bereitgestellt.

Mithilfe von Flugpassagierdaten aus den Jahren 2006-2010 wurde das statistische Modell mittels einer Regressionsanalyse entwickelt. Die einzelnen Forecasts des Algorithmus für die 50 U.S.-Bundesstaaten sind in Anhang 2 dargestellt. Der Algorithmus hat eine Genauigkeit von 4,32 im Mittelwert bei einem Median von 3,0 und einer Standardabweichung von 4,22. Die größte Abweichung beträgt 17. Abbildung 27 gibt einen Überblick zur Forecastinggenauigkeit des Algorithmus.

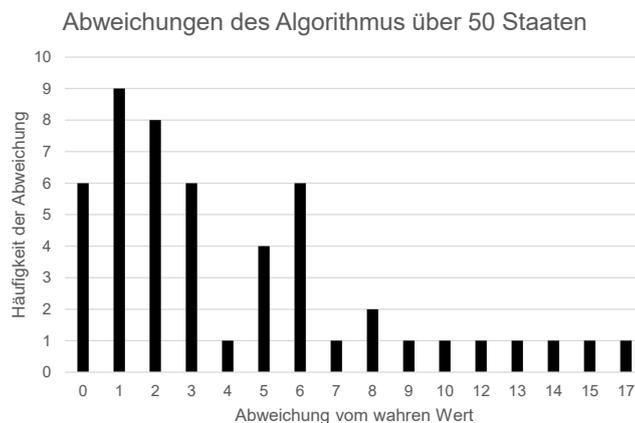


Abbildung 27: Forecastinggenauigkeit des Algorithmus<sup>389</sup>

Bei der in Tabelle 2 angegebenen Bezahlregel liegen damit 80 % der Forecasts des Algorithmus innerhalb des Bereichs einer Bonuszahlung. Das heißt auch, dass die Forecastinggenauigkeit des Algorithmus mit  $r = 0,877$ ,  $p < 0,001$  hochkorreliert mit der Bezahlregel ist. Basierend auf der Bezahlregel kann durch den Algorithmus ein mittlerer Bonus von 0,49 \$ (0,36 \$) bei einem Median von 0,55 \$ erreicht werden.

### 3.2.4 Abgefragte Variablen

Der Fokus auf die Algorithmen-Aversion i. e. S. definiert die im Experiment abzufragenden Variablen. Da der daraus erwachsende Erkenntnisfortschritt möglichst unmittelbar hervorgehen soll, wird hier auf das Sparsamkeitsprinzip Wert gelegt. Das Sparsamkeitsprinzip – oder auch *Ockhams Rasiermesser* – besagt, dass unnötig komplizierte Erklärungen eines theoretischen Modells *abzurasiieren* sind, damit deren Verständnis und Weiterentwicklung hohe Validität besitzt.<sup>390</sup> In diesem Sinne gilt für das

<sup>389</sup> Eigene Darstellung.

<sup>390</sup> Vgl. Döring und Bortz 2016, S. 57.

theoretische Modell und damit für die abzufragenden Variablen, „[...] mit möglichst wenigen Konstrukten, Axiomen und Theoremen aus[zu]kommen und nur so komplex wie unbedingt notwendig [zu] sein“<sup>391</sup>. Unter diesen Prämissen wird auf die Variablen fokussiert, die die Algorithmen-Aversion erklären. Diese sind die Confidence in den eigenen Forecast und die Confidence in den Forecast des Algorithmus (siehe Tabelle 3). Diese wurden in randomisierter Reihenfolge gestellt. Für die Untersuchung zu RPI wird den Teilnehmern der Studie erklärt, dass ihre Peers zwischen dem eigenen Forecast und dem eines Algorithmus wählen durften.<sup>392</sup> Da es bislang keine Untersuchung gibt, die einen ähnlichen Aufbau hat, wird in Studie 2 eine explorative Frage implementiert (Tabelle 3, Frage 3). Diese Variable fragt ab, wie ausgeprägt ein Teilnehmer die Algorithmen-Aversion bei seinen Peers einschätzt.

Tabelle 3: Abgefragte Variablen im Experiment

Num- ber	Question	Scale
1	How much confidence do you have in the statistical model's estimates?	1 = none; 5 = a lot
2	How much confidence do you have in your estimates?	1 = none; 5 = a lot
3 <sup>393</sup>	What do you think, how many of the other MTurkers chose to tie their incentive to the model's accuracy?	0–100 %

Um möglichst wenig am Ablauf des Experiments zu verändern, werden einige weitere Fragen aus Dietvorst et al. (2015) implementiert. Diese haben jedoch keinen Erklärungsgehalt für die Zusammenhänge in der Algorithmen-Aversion und folgen demnach auch keinem weiteren Zweck in dieser Untersuchung. Sie werden nicht näher betrachtet.<sup>394</sup> Abschließend werden die Teilnehmer nach ihrem Alter, ihrem Geschlecht und ihrem höchsten Bildungsabschluss befragt.

<sup>391</sup> Döring und Bortz 2016, S. 57.

<sup>392</sup> Näheres dazu in Abschnitt 5.3.1.

<sup>393</sup> Nur in Studie 2 abgefragt. Siehe hierzu Kapitel 5.

<sup>394</sup> Diese Fragen sind: „How likely is the model to make a really bad estimate?“, „Do you think that your estimate or the model's estimate is closer to the last state's true rank?“ und „On average, how many ranks do you think your estimates are away from states' actual ranks?“. Siehe Anhang 6.

### 3.2.5 Besonderheiten durch die Replikation

Im Folgenden wird beschrieben, welche weiteren Veränderungen am experimentellen Aufbau nach Dietvorst et al. (2015) vorgenommen werden, welche nicht auf die Treatments zurückzuführen sind. Auf der Startseite des Experiments wird kurz eingeführt, welche Institution die Studie durchführt. Hier wird entsprechend die Universität Ulm anstelle der University of Pennsylvania eingefügt. Weiter wird nach einer ersten Beschreibung der anstehenden Aufgabe eine Seite eingefügt, in der abgefragt wird, ob die Teilnehmer diese oder eine ähnliche Aufgabe in der Vergangenheit bereits durchgeführt haben. Teilnehmer, die diese Frage bejaht haben, konnten das Experiment nicht weiter durchführen. Um den hier herausfallenden Teilnehmern möglichst wenig Zeit zu nehmen, wird diese Frage so früh wie möglich im Experiment gestellt. Für die Kontrollgruppen und damit die Replikation sind dies bis zur Entscheidung im incentivierten Forecast die einzigen Änderungen gegenüber dem Experimentalsetting von Dietvorst et al. (2015). Nach der Entscheidung im incentivierten Forecast wurden einige Fragen entfernt, die Dietvorst et al. (2015) als explorative Fragen gestellt und die keine nennenswerten Erkenntnisse hervorgebracht haben. Die letzte Seite dient der Vermittlung des Codes zur Bezahlung der Teilnehmer und einer Danksagung für die Teilnahme. Diese wurde um den Hinweis ergänzt: „We are very aware of your performance and we will pay your bonus as quickly as possible“. Bonuszahlungen sind ein nicht durch MTurk kontrolliertes Instrument und bedürfen daher erhöhter Seriosität hinsichtlich der Sicherheit der Auszahlung. Dem wurde hier Rechnung getragen.

### 3.3 Methodik zur Datenauswertung

Die Datenauswertung und die Hypothesentests erfolgen durch IBM SPSS Statistics 28. Um die Vergleichbarkeit der Ergebnisse mit den bestehenden Erkenntnissen der Algorithmen-Aversion zu gewährleisten, orientiert sich die Auswertung der Daten eng an der bestehenden Literatur. Das bedeutet, die Hypothesentests werden mittels Chi-Quadrat-Tests durchgeführt. Weiter sind für den Vergleich der Performancemaße (Forecastinggenauigkeit und Bonuszahlungen) und der Confidencewerte (in den eigenen Forecast und in den Forecast des Algorithmus) t-Tests von Relevanz. Dabei ist hervorzuheben, dass die Daten zu den Confidencewerten auf einer Likert-Skala erfasst wurden. Analog zur bestehenden Literatur ist eine Äquidistanz zwischen den

Skalenpunkten gegeben.<sup>395</sup> Die Ergebnisse können demnach als metrisch betrachtet und interpretiert werden.<sup>396</sup> Neben der Analyse zum Vergleich von Gruppen ist es im Untersuchungskontext notwendig, die Mediationsanalyse zu verstehen. Die von Dietvorst et al. (2015) gezeigten Zusammenhänge der Confidence als Mediator für die Algorithmen-Aversion werden repliziert und erweitert. Hierzu wird eine Mediationsanalyse mittels Bootstrapping durchgeführt. Für alle statistischen Tests wird in der vorliegenden Arbeit ein empirisches Signifikanzniveau von Alpha = 5 % angenommen. Dieses definiert die Irrtumswahrscheinlichkeit bei der Ablehnung der Nullhypothese und entspricht mit 5 % dem geläufigsten Wert.<sup>397</sup> Entsprechend ist ein p-Wert kleiner 5 % für die Annahme der Hypothesen definiert.

Der Chi-Quadrat-Test ( $\chi^2$ ) dient der Überprüfung zweier Variablen auf ihre stochastische Unabhängigkeit.<sup>398</sup> Er basiert auf Kreuztabellen. Kreuztabellen erfassen die Häufigkeiten bzw. die Häufigkeitsverteilung zweier Variablen.<sup>399</sup> Diese Grundform macht die Chi-Quadrat-Analyse zu einer flexiblen und umfangreichen Analysemethode für Signifikanztests. Es lassen sich nominal-, ordinalskalierte und (gruppierte) metrische Variablen analysieren. Durch die Darstellung in Kreuztabellen werden zunächst univariate Häufigkeiten in Zeilen und Spalten gezählt. Bei gerichteten Hypothesen erfolgt die Darstellung der abhängigen Variablen über die Zeilen. Die beobachteten Häufigkeiten ( $n_{ij}$ ) werden in der Kreuztabelle dargestellt und mit den bei Unabhängigkeit erwarteten Häufigkeiten  $\tilde{n}_{jl}$  verglichen:

$$\tilde{n}_{jl} = \frac{n_j \cdot n_l}{n}$$

Die Berechnung des  $\chi^2$  erfolgt über L-Spalten und J-Zeilen. Je mehr sich  $n_{jl}$  und  $\tilde{n}_{jl}$  unterscheiden, desto größer ist das  $\chi^2$ .<sup>400</sup> Je größer wiederum das  $\chi^2$  ausfällt, desto größer ist die Unabhängigkeit der untersuchten Variablen:

$$\chi^2 = \sum_{j=1}^J \sum_{l=1}^L \frac{(n_{jl} - \tilde{n}_{jl})^2}{\tilde{n}_{jl}}$$

---

<sup>395</sup> Siehe beispielsweise Dietvorst et al. 2015, S. 120.

<sup>396</sup> Vgl. Kromrey et al. 2016, S. 239.

<sup>397</sup> Vgl. Backhaus et al. 2021, S. 36.

<sup>398</sup> Vgl. Döring und Bortz 2016, S. 239; Sibbertsen und Lehne 2021, S. 391.

<sup>399</sup> Vgl. Kuckartz et al. 2010, S. 82.

<sup>400</sup> Vgl. Sibbertsen und Lehne 2021, S. 101.

Dabei gilt, dass  $n_{ij}$ , also jede Zelle, mindestens fünf Datenpunkte benötigt.<sup>401</sup> Eine Ausnahme davon gilt, wenn eine Kreuztabelle größer 2x2 vorliegt und dabei mehr als 20 % der Zellen mehr als fünf Datenpunkte aufweisen sowie jede Zelle mindestens den Wert 1 aufweist.<sup>402</sup> Wenn dem nicht so ist, können entweder Zellen zusammengefasst werden, also beispielsweise Altersgruppen eingeführt werden, oder es kann alternativ der Exakte Test nach Fisher gerechnet werden.

Grundsätzlich testet das Chi-Quadrat zweiseitig, das heißt, es wird auf einen Unterschied der beiden Variablen getestet, jedoch nicht, ob der Unterschied positiv oder negativ ausfällt. In der vorliegenden Arbeit werden die Hypothesen theoriebasiert so hegeleitet, dass sie eine eindeutige Richtung der Beeinflussung der Algorithmen-Aversion prognostizieren.<sup>403</sup> Das heißt, das Chi-Quadrat sollte einseitig getestet werden. Dazu ist anzumerken, dass einseitiges Testen nur dann zulässig ist, wenn eine 2x2-Kreuztabelle vorliegt, denn hier entspricht die Chi-Quadrat-Verteilung, mit nur einem Freiheitsgrad, der quadrierten Normalverteilung.<sup>404</sup> Für weitere Chi-Quadrat-Tests in den Zusatzanalysen dieser Untersuchung, für die keine Hypothesen aufgestellt werden, wird zweiseitig getestet.

Für signifikante Chi-Quadrat-Tests wird die jeweilige Effektstärke durch Cramer's V berichtet. Cohen (2013) spricht bei  $V = 0,1$  von einem schwachen Effekt, bei  $V = 0,3$  von einem mittleren Effekt und bei  $V = 0,5$  von einem starken Effekt.<sup>405</sup> Die zu erwartenden Effekte werden basierend auf den Daten von Dietvorst et al. (2015) errechnet: Bei einem Chi-Quadrat-Test von  $\chi^2 = 8,478$  mit  $p = 0,004$  und  $n = 522$  errechnet sich ein Cramer's -V von 0,127. Das heißt, es liegt für die Algorithmen-Aversion grundsätzlich ein kleiner Effekt vor. Demnach sind Beeinflussungen dieses Effekts, wie sie in dieser Arbeit untersucht werden, ebenfalls als kleiner Effekt zu erwarten. In Dietvorst et al. (2015) wird dieser kleine Effekt mit einer Gruppengröße von rund neunzig Teilnehmern gewährleistet. Um gesicherte, nicht zufallsbasierte Ergebnisse zu erhalten, werden die größten Gruppengrößen aus Dietvorst et al. (2015) adaptiert – dies ist Studie 3b mit rund 250 Teilnehmern je Gruppe.

---

<sup>401</sup> Vgl. Sibbertsen und Lehne 2021, S. 393.

<sup>402</sup> Vgl. Yates et al. 1999, S. 734.

<sup>403</sup> Hypothesen sind demnach: "[...] wissenschaftliche begründete Vermutungen über einen Tatbestand oder über einen Zusammenhang von mindestens zwei Merkmalen [...]". Vgl. Häder 2015, S. 33.

<sup>404</sup> Die Freiheitsgrade errechnen sich mit  $v = (J-1)(L-1)$ .

<sup>405</sup> Vgl. Cohen 2013, S. 227.

Der Mittelwert ist das relevanteste deskriptive Maß und lässt eine zuverlässige Charakterisierung quantitativer Daten zu.<sup>406</sup> Er kann grundsätzlich als Lagemaß verstanden werden. Bei metrischen Variablen stellt er dabei den zentralen Wert dar. Je symmetrischer die zugrundeliegenden Daten verteilt sind, desto aussagekräftiger ist der Mittelwert. Der Mittelwert hat eine Zentrierungseigenschaft. Das heißt, dass die Summe der positiven und negativen Abweichungen immer Null ergibt. Zur Beurteilung eines Mittelwerts wird außerdem die Standardabweichung herangezogen. Als Quadratwurzel der Varianz hilft die Standardabweichung in der Beurteilung der Verteilung der Daten um den Mittelwert.<sup>407</sup>

Um signifikante Differenzen eines oder mehrerer Mittelwerte zu testen, wird ein t-Test durchgeführt. Voraussetzung für einen t-Test sind stetige Daten, eine zufällige Stichprobe und eine annähernde Normalverteilung.<sup>408</sup> Bei Stichproben, die  $n > 30$  sind, kann die Normalverteilungsannahme vernachlässigt werden. Es muss zunächst unterschieden werden, ob ein Einstichprobentest oder ein Zweistichprobentest zugrunde gelegt wird. Beim Einstichprobentest wird untersucht, ob ein Mittelwert von einem bestimmten gegebenen Wert abweicht. Dies ist in der vorliegenden Untersuchung nicht von Relevanz. Für die folgende Beurteilung der empirischen Erhebung sind Zweistichprobentests relevant. Beim Zweistichprobentest ist weiter zu unterscheiden, ob die Stichproben verbunden oder unverbunden sind. Für verbundene Stichproben wird im Folgenden vom Paardifferenzentest gesprochen. Er wird angewendet, wenn beispielsweise die Leistung des Algorithmus mit der Leistung der Forecaster innerhalb einer Experimentalgruppe auf Unterschiede hin untersucht werden. Werden hingegen die Mittelwerte einer Kontrollgruppe und einer Treatmentgruppe auf Unterschiede im Mittelwert der Leistung untersucht, wird vom ungepaarten t-Test gesprochen. Voraussetzung dieses Tests ist die Varianzgleichheit der zu vergleichenden Gruppen. Ist diese nicht gegeben, wird auf den Welch-Test ausgewichen.<sup>409</sup>

Für eine zuverlässige Bewertung der Ergebnisse dieser Untersuchung und eine weitreichende Erklärung der Zusammenhänge der Algorithmen-Aversion ist es notwendig, die bekannten empirischen Zusammenhänge zu berücksichtigen und darauf aufzubauen. Das heißt insbesondere, dass bei signifikanten Unterschieden im Chi-

---

<sup>406</sup> Vgl. Backhaus et al. 2021, S. 19–20.

<sup>407</sup> Vgl. Backhaus et al. 2021, S. 21.

<sup>408</sup> Vgl. Döring und Bortz 2016, S. 250.

<sup>409</sup> Siehe dazu auch Rasch et al. 2011.

Quadrat-Test zur Algorithmen-Aversion bei gleichzeitigen signifikanten Differenzen in den Mittelwerten der Confidencewerte (Confidence in den eigenen Forecast und Confidence in den Forecast des Algorithmus) eine Mediationsanalyse durchgeführt wird. Diese kann zeigen, ob gilt, dass die abnehmende Confidence in den Forecast des Algorithmus ein Mediator für die Algorithmen-Aversion ist. Dieser Effekt wurde bereits über mehrere Experimente nachgewiesen.<sup>410</sup> Mediationen zu untersuchen, ist insbesondere dann relevant, wenn solche robusten Beziehungen zwischen unabhängigen und abhängigen Variablen bestehen.<sup>411</sup> Mediationsanalysen werden verwendet, um Hypothesen über die Wirkung von unabhängigen auf abhängige Variablen zu testen bzw. tiefgründiger zu verstehen.<sup>412</sup> Die Mediationsanalysen in dieser Arbeit werden mittels des PROCESS Makro von Andrew F. Hayes in IBM SPSS Statistics 28 durchgeführt.<sup>413</sup> Dieses Programm führt eine lineare Regression zur Mediationsanalyse durch. Der Mediationseffekt wird mittels Bootstrapping geschätzt. Die Grundlagen für diese Mediationsanalyse werden im Folgenden gezeigt.

Die schlichteste Beziehungsform stellt der Effekt einer Ursache bzw. unabhängigen Variable [X] auf ein bestimmtes Ergebnis bzw. eine abhängige Variable [Y] dar.<sup>414</sup> Diese Beziehung kann durch eine Vielzahl an Variablen ermöglicht, also mediiert werden.<sup>415</sup> Vor allem in der Psychologie werden Mediationsanalysen durchgeführt, um indirekte Effekte zu ermitteln. Dabei sei beispielsweise die Wirkung von Motivation auf eine Schulnote betrachtet. Mediiert werden kann die Wirkung von Motivation auf eine Schulnote durch den indirekten Effekt der Lerndauer. Im Kontext der Algorithmen-Aversion führt das Leistungsfeedback zum Algorithmus zu einer Algorithmen-Aversion. Diese Aversion wird mediiert durch die Reduktion der Confidence in den Algorithmus, wohingegen das Leistungsfeedback zum eigenen Forecast keinen Einfluss auf die Algorithmen-Aversion hat.<sup>416</sup> Es ist daher bei Mediationsanalysen bedeutsam zu verstehen, welcher Prozess für die analysierten Effekte verantwortlich ist.<sup>417</sup>

---

<sup>410</sup> Siehe hierzu Abschnitt 4.4.3.1 und 5.4.2.3 sowie Dietvorst et al. 2015, S. 121.

<sup>411</sup> Vgl. Holmbeck 1997, S. 599.

<sup>412</sup> Vgl. Hayes und Rockwood 2017, S. 40.

<sup>413</sup> Die statistische Auswertung mittels PROCESS ist in der Controlling-Forschung breit etabliert. Siehe dazu beispielsweise: Bisbe und Malagueño 2012; Fleischman et al. 2017.

<sup>414</sup> Vgl. Preacher und Hayes 2004, S. 717.

<sup>415</sup> Vgl. Baron und Kenny 1986, S. 1176; Holmbeck 1997, S. 599,

<sup>416</sup> Vgl. Dietvorst et al. 2015, S. 121.

<sup>417</sup> Vgl. Preacher und Hayes 2004, S. 717.

Preacher et al. (2007) postulieren, dass es sich um eine Mediation bzw. einen indirekten Effekt handelt, wenn ein kausaler Effekt einer unabhängigen Variablen X durch einen Mediator M auf eine abhängige Variable Y übertragen wird.<sup>418</sup>

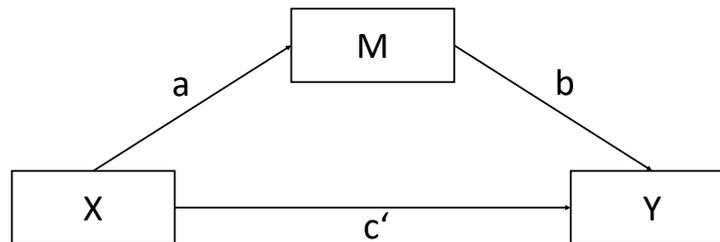


Abbildung 28: Schematische Darstellung der Mediationsanalyse<sup>419</sup>

Sachlogisch kann gesagt werden, dass X auf Y wirkt, weil X auf die Mediatorvariable M wirkt. Die Mediatorvariable wiederum übermittelt die Wirkung von X auf Y, indem M auf Y wirkt. Das bedeutet, dass ein Mediationskonstrukt als eine Folge von zwei oder mehreren kausalen Geschehnissen verstanden werden kann, die in einer Abfolge von  $X \rightarrow M \rightarrow Y$  erfolgt. Aufgrund dieser Tatsache muss das zeitliche Auftreten eines Mediators zwischen dem Auftreten von X und Y liegen.<sup>420</sup> Das bedeutet, die Interpretation einer Mediation sollte entlang dieser Pfade vorgenommen werden. Die Pfade in Abbildung 28 können dabei wie folgt beschrieben werden:<sup>421</sup>

Der Pfad  $c'$  spiegelt den direkten Effekt von X auf Y unter der Kontrolle von M wider.<sup>422</sup> Das Vorzeichen, das  $c'$  trägt, gibt an, ob die Erhöhung einer Einheit der unabhängigen Variable die abhängige Variable ebenfalls erhöht ( $c' = +$ ) oder senkt ( $c' = -$ ).<sup>423</sup> Analog dazu verhält sich die Wirkung der Vorzeichen von Pfad a bzw. b auf M bzw. Y. Pfad a stellt dabei den Koeffizienten von X für ein Modell dar, das M durch X vorhersagt. Pfad b und  $c'$  wiederum sind Koeffizienten in einem Modell, das Y durch M bzw. durch X vorhersagt. Ist sowohl der Koeffizient für a als auch für b signifikant, kann von einer Mediation gesprochen werden. Das Produkt von Pfad a und b beschreibt den indirekten Effekt, also die Mediation. Können alle drei Variablen beobachtet werden, dann kann der Gesamteffekt ermittelt werden. Der Gesamteffekt lautet:

<sup>418</sup> Vgl. Preacher et al. 2007, S. 185.

<sup>419</sup> Eigene Darstellung nach Hayes 2018, S. 83.

<sup>420</sup> Vgl. Hayes und Rockwood 2017, S. 40.

<sup>421</sup> Für eine ausführliche mathematische Beschreibung aller Pfade siehe Hayes 2018, S. 82–86.

<sup>422</sup> Vgl. Preacher und Hayes 2004, S. 717

<sup>423</sup> Bei einer binären Variablen bezieht sich dieser Zusammenhang auf den Gruppenmittelwert. Siehe Hayes 2018, S. 83.

$c = c' + ab$ . Anzumerken ist, dass in Modellen mit binären Mediatorvariablen oder binären abhängigen Variablen dieses Kalkulationsschemata nicht immer zutrifft.<sup>424</sup> Der direkte Effekt kann daher in der vorliegenden Arbeit nicht über die Mediationsanalyse bestimmt werden. Hier wird auf eingangs dieses Abschnitts erwähnte Untersuchung der Algorithmen-Aversion mittels Chi-Quadrat-Test verwiesen.

Eine Mediation, bei der der direkte und der indirekte Effekt nicht das gleiche Vorzeichen tragen, wird als inkonsistente Mediation bezeichnet.<sup>425</sup> In der Betrachtung mehrerer Mediatoren können diese unterschiedliche Vorzeichen und unterschiedlich große Effekte aufweisen. Das heißt, unterschiedliche Mediatoren können den Haupteffekt verstärken oder abschwächen. Es ist daher für das logische Verständnis einer Mediationsanalyse notwendig, die erwarteten Zusammenhänge im Vorfeld theoretisch abzuleiten.<sup>426</sup>

Die Signifikanz des indirekten Effekts wird in PROCESS mittels Bootstrapping getestet.<sup>427</sup> Bootstrapping führt ein Resampling des zugrundeliegenden Datensatzes durch. Das heißt, es wird aus der erhobenen Stichprobe wiederholt und mit Zurücklegen eine zufällige neue Stichprobe gezogen. Somit kann jeder Datenpunkt erneut gezogen werden, sobald eine neue Stichprobe der Größe  $n$  gebildet wird – es kann im Umkehrschluss jedoch auch vorkommen, dass ein Datenpunkt nicht in der Berechnung berücksichtigt ist. Nachdem neue Stichprobenziehungen erzeugt wurden, werden die Pfade  $a$  und  $b$  wie auch das Produkt der Koeffizienten von  $a$  und  $b$  geschätzt und festgehalten. Dabei wird der Vorgang insgesamt  $k$ -mal wiederholt. Hayes (2009) spricht von mindestens 1000 Iterationen – im Optimalfall von 5000 Iterationen.<sup>428</sup> In der vorliegenden Arbeit wird dieser Empfehlung gefolgt und es werden entsprechend 5000 Bootstrap Samples gebildet. Nach der Berechnung liegen demnach  $k = 5000$  Schätzungen des indirekten Effekts vor. Deren Verteilung dient als empirische Annäherung an die Stichprobenverteilung des indirekten Effekts. Dadurch können Schlussfolgerungen über die Größe des indirekten Effekts der Grundgesamtheit gezogen werden. Dabei werden die  $k$ -Schätzungen verwendet, um ein %- Konfidenzintervall zu generieren. Das Konfidenzintervall wird erzeugt, indem die  $k$ -Werte des indirekten Effekts

---

<sup>424</sup> Vgl. Hayes 2018, S. 545, 2009, S. 409.

<sup>425</sup> Vgl. MacKinnon et al. 2007, S. 602.

<sup>426</sup> Vgl. Hayes 2018, S. 81.

<sup>427</sup> Vgl. Hayes 2009, S. 412.

<sup>428</sup> Vgl. Hayes 2018, S. 98, 2009, S. 412.

aufsteigend angeordnet werden. Anhand der geordneten Datenmenge werden die Unter- und Obergrenze eines Konfidenzintervalls festgelegt. Aus diesem Verfahren resultiert ein Perzentil basiertes Bootstrap-Konfidenzintervall. Wenn dabei Null nicht im Konfidenzintervall enthalten ist, kann gezeigt werden, dass der indirekte Effekt von Null verschieden ist.<sup>429</sup> Ein indirekter Effekt wird als signifikant angenommen, wenn das Konfidenzintervall die Null nicht einschließt.

Bootstrapping ist damit eine valide und leistungsfähige Methode zur Prüfung von Mediationseffekten.<sup>430</sup> Durch Bootstrapping kann eine hohe statistische Power sichergestellt und der Fehler erster Art weitestgehend ausgeschlossen werden.<sup>431</sup> Darüber hinaus ist Bootstrapping unabhängig von den Verteilungseigenschaften der zugrunde liegenden Daten.<sup>432</sup> Eine zentrale Voraussetzung für das Bootstrapping ist neben der theoretischen Fundierung aller Zusammenhänge und der Signifikanzen der einzelnen Pfade die Homoskedastizität. PROCESS hat dabei die Möglichkeit, heteroskedastizität-konsistente Standardfehler (HC-Standardfehler) einzusetzen.<sup>433</sup> Es wird daher in dieser Arbeit standardmäßig der von Hayes und Cai (2007) empfohlene HC-Standardfehler HC3 nach Davidson und MacKinnon (1993) verwendet.<sup>434</sup>

---

<sup>429</sup> Vgl. Hayes 2009, S. 412.

<sup>430</sup> Vgl. Hayes 2009, S. 412.

<sup>431</sup> Vgl. Hayes 2009, S. 412. Siehe dazu auch die Simulationsstudien von MacKinnon et al. 2002.

<sup>432</sup> Vgl. Preacher et al. 2007, S. 185.

<sup>433</sup> PROCESS implementiert hier Davidson und MacKinnon 1993.

<sup>434</sup> Vgl. Hayes und Cai 2007, S. 716.

## 4 Studie 1

In Studie 1 wird die Wirkung von Zeitdruck, „do your best“-Zielen und den Entscheidungsrechten in der Interaktion mit einem Algorithmus untersucht. Aufbauend auf die Vertiefung der Motivation zu Studie 1 (Abschnitt 4.1) werden in Abschnitt 4.2 die Hypothesen entwickelt. In Abschnitt 4.3 und 4.4 folgen das Experiment und die Hypothesentests. Daran schließen ebenfalls in Abschnitt 4.4 die Schlussfolgerungen und die Limitationen denen Studie 1 unterliegt an. Abbildung 29 zeigt den Aufbau von Kapitel 4. Es ist anzumerken, dass die empirischen Ergebnisse von Studie 1 bereits im „Journal of Management Control“ unter dem Titel „Towards a better understanding on mitigating algorithm aversion in forecasting: an experimental study“<sup>435</sup> publiziert wurden.

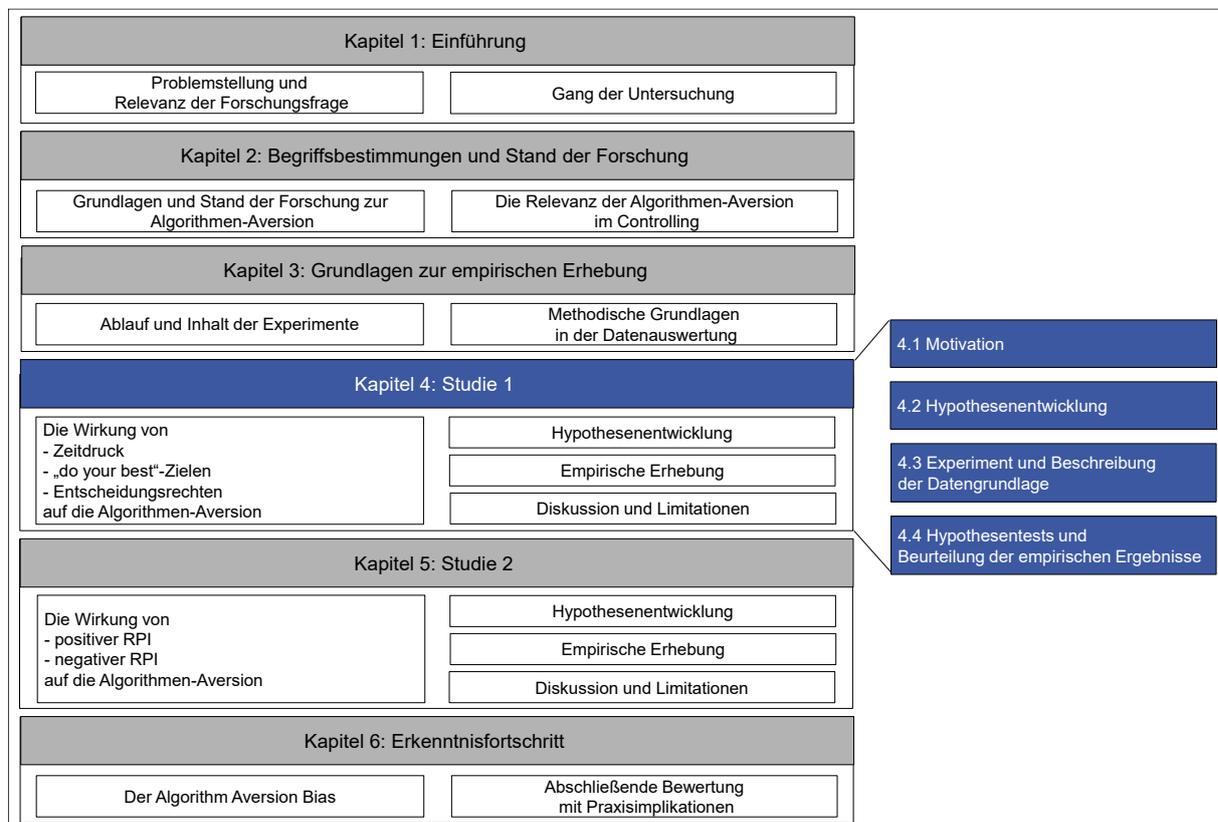


Abbildung 29: Aufbau von Kapitel 4<sup>436</sup>

<sup>435</sup> Vgl. Jung und Seiter 2021.

<sup>436</sup> Eigene Darstellung.

### 4.1 Motivation

Abschnitt 1.2 sowie 2.2.1 zeigen, wie groß das Potenzial und der damit einhergehende Handlungsdruck zum Einsatz von Algorithmen in Forecasts ist. Dabei wurde insbesondere herausgestellt, dass die unternehmerischen Anstrengungen darin bestehen, Algorithmen dauerhaft in betrieblichen Abläufen zu etablieren.<sup>437</sup> Dies führt dazu, dass Forecaster Leistungsfeedback zu den Algorithmen erhalten, mit denen sie zusammenarbeiten. Das Leistungsfeedback, als Grundlage für eine Algorithmen-Aversion, ist demnach künftig omnipräsent (H1). Der Fokus dieser Studie liegt daher auf dem besseren Verständnis der Algorithmen-Aversion im Arbeitsumfeld von Forecastern. Die bisherige Forschung zeigt, dass die Ausprägung der Aversion in Abhängigkeit unterschiedlicher Gegebenheiten schwankt.

Dietvorst et al. (2018) konnten zeigen, dass die Algorithmen-Aversion durch die Möglichkeit des Eingreifens des Forecasters in die Ergebnisse des Algorithmus gesenkt wird.<sup>438</sup> Es spielt keine Rolle, ob der Forecaster das Ergebnis des Algorithmus stark oder nur geringfügig verändern kann; selbst kleine Anpassungen mindern die Algorithmen-Aversion erheblich. Da sich die Forecastinggenauigkeit dadurch verschlechtert, ist dies möglicherweise nicht immer eine praktikable Lösung. Castelo et al. (2019) zeigen, dass die Algorithmen-Aversion besonders ausgeprägt ist, wenn Aufgaben einen subjektiven Charakter (vs. objektiven Charakter) haben. Weitere Studien deuten darauf hin, dass die Anzahl der Interaktionen mit einem Algorithmus die Algorithmen-Aversion beeinflusst.<sup>439</sup>

Die Forschung vernachlässigt bisher jedoch das Arbeitsumfeld von Forecastern. Da die tägliche Arbeit von Forecastern in das Arbeitsumfeld eingebettet ist, muss verstanden werden, wie dieses mit der Algorithmen-Aversion interagiert. Aufbauend auf den Überlegungen in Abschnitt 2.2.2.2 werden dazu in Studie 1 drei Einflussfaktoren des Arbeitsumfelds von Forecastern auf deren Algorithmen-Aversion untersucht. Erstens spielen bei der Erstellung von Forecasts Fristen eine große Rolle. Daten können kurzfristig eintreffen und müssen schnell verarbeitet werden, was in der Regel zu Zeitdruck bei der Erstellung von Forecasts führt (H2). Zweitens sind Zielvorgaben in

---

<sup>437</sup> Zur Einführung von Algorithmen in Forecasts siehe Abschnitt 2.2.1 oder auch Appelbaum et al. 2017, S. 29–44; Bergmann et al. 2020, S. 29; Seiter 2019, S. 55.

<sup>438</sup> Siehe auch Carbone et al. 1983; Goodwin und Fildes 1999.

<sup>439</sup> Vgl. Feng und Gao 2020, o. S.; Filiz et al. 2021, o. S. Für weitere Einflussfaktoren auf die Algorithmen-Aversion siehe Abschnitt 2.1.5.

der Praxis zwar spezifisch, jedoch entsprechen sie insbesondere für einzelne Forecasts eher einem „do your best“-Ziel, an das keine unmittelbare leistungsabhängige Vergütung gekoppelt ist (H3). Drittens sind Forecasts eine wiederkehrende Aufgabe und die Forecaster kennen die Datengrundlage für den Forecast. Sie können daher beurteilen, inwieweit sie für die Erstellung eines Forecasts geeignet sind, und häufig mitentscheiden, mit welchen Daten die Algorithmen arbeiten sollen (H4). Die Verknüpfung dieser Einflussfaktoren auf die Algorithmen-Aversion stützt sich auf das Framework von Bonner und Sprinkle (2002) zur Untersuchung Arbeitsumfelds, die auf einen Forecaster wirken (siehe Abschnitt 2.2.2). Im Folgenden werden die Variablen des Arbeitsumfelds von Forecastern näher beschrieben und gezeigt, welche Auswirkungen diese auf die Algorithmen-Aversion haben können. Dem vorangestellt wird zunächst die Algorithmen-Aversion selbst untersucht und nachgewiesen.

## 4.2 Hypothesenentwicklung

### 4.2.1 Algorithmen-Aversion

Die Algorithmen-Aversion beruht auf der Reduktion der Confidence in einen Algorithmus, wenn die Leistung des Algorithmus wahrgenommen wird.<sup>440</sup> Confidence beschreibt die Erwartungen und Zuversicht des Forecaster hinsichtlich der Genauigkeit eines Forecasts.<sup>441</sup> Die anfänglichen Erwartungen eines Forecasters in Bezug auf einen Algorithmus können durch vergangene Erfahrungen getrieben sein und sind daher im Zeitverlauf instabil. Die Interaktion mit Algorithmen über viele Jahre im Alltag sowie im Berufsleben und die kontinuierliche Verbesserung der Algorithmen kann zu vermehrten positiven Erfahrungen mit dem Ergebnis von Algorithmen und damit zur Confidence in unbekannte Algorithmen führen.<sup>442</sup> Daraus folgt, dass die Algorithmen-Appreciation ein über die Zeit immer robuster werdendes Phänomen darstellt.

Unternehmen fördern das kontinuierliche Zusammenarbeiten mit Algorithmen in Forecasts. Das hieraus resultierende Leistungsfeedback führt zu einer Algorithmen-Aversion. Es ist daher weniger relevant, wie die Algorithmen-Appreciation sich in Zukunft verhält. Stattdessen wird es zunehmend bedeutsamer, die Algorithmen-Aversion besser zu verstehen. Aus einem ansteigenden Ausgangsniveau der Algorithmen-Appreciation

---

<sup>440</sup> Siehe Abschnitt 2.1.3.

<sup>441</sup> Siehe Abschnitt 2.1.3 und 2.1.4.3.

<sup>442</sup> Vgl. Al-Htaybat und Alberti-Alhtaybat 2017, S. 850; Quattrone 2016, S. 118.

## Studie 1

preciation resultiert jedoch ein potenziell größerer Verlust der Confidence in einen Forecast durch Leistungsfeedback. Das negative Potenzial für die Algorithmen-Aversion steigt demnach mit steigender Algorithmen-Appreciation an. Dies führt zu steigender Algorithmen-Aversion. Konkret bedeutet das, dass über die Zeit der Gesamteinsatz von Algorithmen steigen wird, jedoch das Delta der Algorithmen-Aversion, vor und nach Leistungsfeedback, weiter anwachsen wird (siehe dazu die exemplarische Darstellung in Abbildung 30). Es wird daher zunächst die Hypothese zum Vorliegen einer Algorithmen-Aversion aufgestellt und überprüft:

H1: Durch Leistungsfeedback entsteht eine Algorithmen-Aversion.

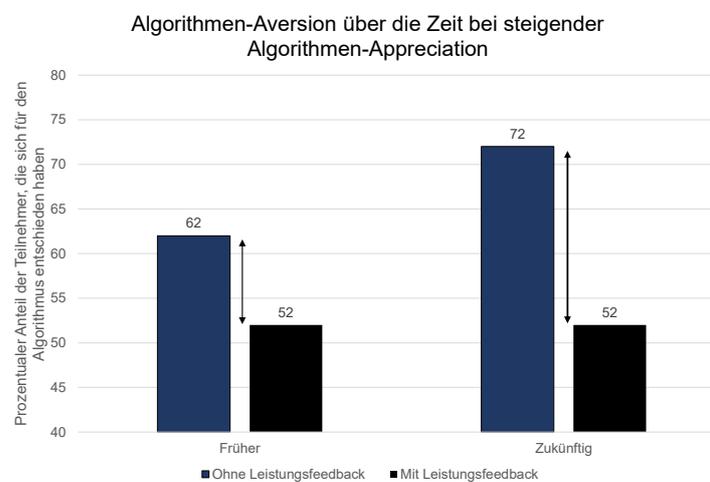


Abbildung 30: Potenzielle Entwicklung der Algorithmen-Aversion durch eine steigende Algorithmen-Appreciation<sup>443</sup>

### 4.2.2 Der Einfluss von Zeitdruck auf die Algorithmen-Aversion

Die Verfügbarkeit von Daten nimmt sowohl in Bezug auf den Umfang als auch auf die Geschwindigkeit ständig zu.<sup>444</sup> In einem schnelllebigen und volatilen Umfeld bleibt Unternehmen und Forecastern immer weniger Zeit, die schwer zu überblickenden Datenmengen in präzise Forecasts zu überführen, wodurch Zeitdruck bei der Erstellung von Forecasts entsteht.<sup>445</sup> Zeitdruck kann dabei aus vielerlei Gründen hervorgehen. Während der Zeitdruck aus einem großen Prozessaufwand der Forecastingaufgabe an sich resultieren kann, kann eine Reduzierung der für eine Aufgabe zur Verfügung stehenden Zeit auch durch Vorgesetzte herbeigeführt werden. So können beispielsweise

<sup>443</sup> Eigene Darstellung.

<sup>444</sup> Vgl. Al-Htaybat und Alberti-Alhtaybat 2017, S. 850.

<sup>445</sup> Vgl. Camerer et al. 2004, S. 862; Spiliopoulos und Ortmann 2018, S. 383.

Fristen verkürzt oder Daten verspätet kommuniziert bzw. nachgereicht werden. Dies kann wiederum beabsichtigt oder aufgrund schlechter zeitlicher Planung entstehen.<sup>446</sup>

Im Folgenden wird gezeigt, welcher Einfluss durch Zeitdruck auf die Erstellung von Forecasts zu erwarten ist und welche Auswirkungen sich auf die Algorithmen-Aversion ergeben. Basierend auf der Analyse zum Stand der Forschung gibt es bislang keine Untersuchung, die Zeitdruck im Kontext einer Interaktion mit einem Algorithmus betrachtet. In der bestehenden Literatur zur Algorithmen-Aversion haben die Forecaster keinen Zeitdruck für die Erstellung der Forecasts. Die Bedeutung und die Auswirkungen von Zeitdruck werden in der Praxis breit diskutiert.<sup>447</sup> Für die vorliegende Untersuchung müssen diese Auswirkungen differenziert betrachtet werden:

Zeitdruck wird von Forecastern als eine der belastendsten Arbeitsanforderungen wahrgenommen.<sup>448</sup> Die Verhaltensforschung zeigt, dass Zeitdruck dazu führt, dass sich Forecaster ängstlich fühlen, das Vertrauen in ihr eigenes Urteilsvermögen reduzieren und teilweise ihre Bemühungen, eine Aufgabe zu erfüllen, verringern.<sup>449</sup> Diese Erkenntnisse finden sich auch in der Literatur zum Thema *choking under pressure* wieder, in der festgestellt wird, dass sich die Leistung von Forecastern verschlechtert, wenn sie aufgrund von Druck einer Aufgabe das Vertrauen in ihre eigenen Fähigkeiten verlieren.<sup>450</sup> Darauf aufbauend neigen Forecaster dazu, aufzugeben, wenn die Leistungserwartungen sehr hoch sind und sie vermindert an sich selbst glauben.<sup>451</sup> Die Folgen von Zeitdruck können kurzfristiger Stress oder langfristige Erscheinungen wie Erschöpfung, Unzufriedenheit oder psychosomatische Beschwerden sein.<sup>452</sup> Begleiterscheinungen dieser Beschwerden sind häufig hohe Ausfallzeiten und eine steigende Fluktuation.<sup>453</sup> Kirchler et al. (2017) empfehlen daher Zeitdruck in bedeutsamen Entscheidungssituationen grundsätzlich auf ein Minimum zu reduzieren. Den Zeitdruck zu reduzieren, geht jedoch mit einem gewissen Aufwand die Umsetzung eines systematischen Zeitmanagements einher.<sup>454</sup>

---

<sup>446</sup> Vgl. Mischer et al. 2021, S. 19–23; Nandhakumar und Jones 2001, S. 193.

<sup>447</sup> Vgl. Kelly et al. 2011; Lambert et al. 2017; Pietsch und Messier 2017; Spilker 1995; Wegier und Spaniol 2015.

<sup>448</sup> Vgl. Mischer et al. 2021, S. 19.

<sup>449</sup> Vgl. Pietsch und Messier 2017, S. 51.

<sup>450</sup> Vgl. Beilock und Carr 2001, S. 701.

<sup>451</sup> Vgl. Dai et al. 2018, S. 1667.

<sup>452</sup> Vgl. Collins und Killough 1992, S. 535; Mischer et al. 2021, S. 19.

<sup>453</sup> Vgl. Cavanaugh et al. 2000, S. 70.

<sup>454</sup> Vgl. Kirchler et al. 2017, S. 57. Für weiterführende Literatur zu Methoden zur Reduktion von Zeitdruck siehe Mischer et al. 2021, S. 19.

## Studie 1

Wenn im beruflichen Umfeld Zeitdruck auftritt oder induziert wird, bleibt dem Forecaster weniger Zeit, verschiedene Optionen und Berechnungen für einen Forecast sorgfältig abzuwägen. Zeitdruck als solcher wird dabei nur als Druck empfunden, wenn die zugrunde liegende Entscheidungssituation beschleunigt werden muss. Dies führt in der Regel zu Veränderungen im Verhalten.<sup>455</sup> Die Literatur, die sich mit den Effekten von Zeitdruck in der Entscheidungsfindung beschäftigt, findet sich in einer Vielzahl von Themenbereichen wieder. Unabhängig davon sind die identifizierten Effekte weitestgehend ähnlich in ihrer Wirkung.

Es ist naheliegend, dass unter Zeitdruck Effekte gezeigt werden, bei denen die Informationsverarbeitung beschleunigt wird und damit eine schnellere Entscheidung herbeigeführt werden kann. Diese schnellere Informationsverarbeitung und Entscheidungsfindung wird als Beschleunigung oder *Acceleration* bezeichnet. So führt die *Acceleration* zwar zu einer schnelleren Entscheidungsfindung, jedoch kann diese Verhaltensänderung eine erhöhte Fehlerquote bewirken und somit in schlechteren Entscheidungen resultieren.<sup>456</sup> Kocher und Sutter (2006) illustrieren in einem Experimentalsetting einen negativen Effekt von steigendem Zeitdruck auf die Entscheidungsgüte und die daraus erzielbaren Incentives.<sup>457</sup> Dieser Effekt wurde in der Literatur mehrfach aufgewiesen und eine Vielzahl an Untersuchungen unterstützt die These der abnehmenden Entscheidungsqualität bzw. der erhöhten Fehlerquote durch Zeitdruck.<sup>458</sup>

Darüber hinaus zeigen Kirchler et al. (2017), dass die Fehleranfälligkeit in Entscheidungssituationen mit steigendem Zeitdruck zunimmt und sich die Risikoneigung der Probanden verändert.<sup>459</sup> Bei Gewinnen erhöht sich eine Aversion gegen riskantes Verhalten; für Verluste steigt die Risikoneigung an. In der Folge von Zeitdruck zeigt Kerstholt (1994), dass die Informationsverarbeitung der Teilnehmer seiner Experimentalsstudien zunimmt und sie analog zum Zeitdruck schnellere Entscheidungen treffen. Die beschleunigte Entscheidungsfindung führt jedoch zunehmend zu schlechten Er-

---

<sup>455</sup> Vgl. Wegier und Spaniol 2015, 13.

<sup>456</sup> Vgl. Ben Zur und Breznitz 1981, S. 90.

<sup>457</sup> Vgl. Kocher und Sutter 2006, S. 375.

<sup>458</sup> Vgl. Kocher et al. 2019, S. 216; Kocher und Sutter 2006, S. 375; Maule et al. 2000, S. 283; Svanström 2016, S. 42.

<sup>459</sup> Vgl. Kirchler et al. 2017, S. 56.

## Studie 1

gebnissen, was besonders schwerwiegend ist, wenn keine sinnvolle Anpassung des Verhaltens an den Zeitdruck stattfindet.<sup>460</sup> In dieser Studie lag als Aufgabe eine Computersimulation zu Grunde und die Teilnehmer verursachten durch ihr stoisches Verhalten zunehmend Systemausfälle. Ergo sind die Teilnehmer der Studie nicht in der Lage, den unter Zeitdruck entstehenden Stress richtig zu verarbeiten und ihr Verhalten anzupassen.<sup>461</sup>

Edland und Svenson (1993) untersuchen den Einfluss von Zeitdruck bzgl. der Entscheidung über die Eignung von jeweils zwei angehenden Studenten für einen Studienplatz. Dabei werden die Entscheider, die die Studenten bewerten, in verschiedene Gruppen mit unterschiedlichen, zeitlichen Restriktionen für die Entscheidung eingeteilt. Es wird beobachtet, dass signifikant weniger der zur Verfügung gestellten Informationen herangezogen wurden, wenn der Zeitdruck zugenommen hat. Die Teilnehmer fokussierten sich auf die aus ihrer Sicht relevantesten Informationen für die Entscheidungsfindung. Die Reduktion der für eine Entscheidung herangezogenen Informationen wurde bereits mehrfach gezeigt.<sup>462</sup> Die Strategie von Forecastern, die vorhandenen Informationen einzugrenzen und diese auf die wesentlichen Informationen zu beschränken, wird als *Filtration* bezeichnet.<sup>463</sup>

Zusammengefasst gilt: Je höher der Zeitdruck in Entscheidungssituationen ist, desto:

- schneller werden Informationen verarbeitet
- weniger Informationen können verarbeitet werden
- schlechter wird die daraus folgende Entscheidung

Unter Zeitdruck reduziert sich die Informationsmenge, die verarbeitet werden kann. Da weniger Informationen verarbeitet werden können, muss im Umkehrschluss die Selektion der bedeutsamen Informationen verbessert werden. Im Trade-off aus besserer Informationsselektion und verringerter Informationsverbreitungsmenge überwiegt in der Regel der negative Effekt und die Entscheidungsqualität sinkt.

Im Hinblick auf die Algorithmen-Aversion sollten diese negativen Auswirkungen auf Forecaster vor allem die Confidence in ihren Forecast betreffen. Wenn ein Algorithmus in eine Entscheidungssituation involviert ist, der durch die zeitliche Restriktion

---

<sup>460</sup> Vgl. Kerstholt 1994, S. 89.

<sup>461</sup> Siehe dazu ebenfalls Dror et al. 1999, S. 714–715

<sup>462</sup> Siehe dazu weiterführend Huber und Kunz 2007; Zakay und Wooler 1984.

<sup>463</sup> Vgl. Ben Zur und Breznitz 1981, S. 90.

nicht tangiert wird, sollte dieser einen probaten Ausweg aus der Drucksituation darstellen. Es wird damit erwartet, dass Zeitdruck die Algorithmen-Aversion abschwächt, da er die Confidence in die eigenen Forecasts verringert und somit der Algorithmus in Relation zum eigenen Forecast aufgewertet wird. Es wird folgende Hypothese aufgestellt:

H2: Zeitdruck reduziert die Algorithmen-Aversion.

### 4.2.3 Der Einfluss eines „do your best“-Ziels auf die Algorithmen-Aversion

In der Literatur zur Algorithmen-Aversion wird in der Regel ein Incentive eingesetzt, das die Forecaster dazu anregt, den bestmöglichen und dabei schwer zu erreichenden Forecast zu erstellen. Dies bedeutet, dass ein spezifisches und schwieriges Ziel verbal zugewiesen wird. Die damit verbundene Anreizstruktur des Incentives lässt für die Forecastinggenauigkeit kaum Fehler zu. In einer realen Arbeitsumgebung von Controllern gibt es häufig keine Arbeitsverträge, die das Gehalt an die Genauigkeit eines einzelnen Forecasts bzw. einer Forecastingaufgabe koppeln. Die Forecaster erhalten zu meist ein festes Gehalt unabhängig von ihrer individuellen oder aufgabenspezifischen Leistung. Die Ziele für einen Forecast werden unabhängig von monetären Anreizen formuliert und können daher im Sinne eines „do your best“-Ziels gesehen werden.<sup>464</sup> Es wird daher im Folgenden diskutiert, wie sich unterschiedliche Zielvorgaben auf die Algorithmen-Aversion auswirken können und warum erwartet wird, dass eine Vorgabe, die dem realen Arbeitsumfeld entspricht, mit einem „do your best“-Ziel, die Algorithmen-Aversion abschwächt.

Einer der am häufigsten nachgewiesenen Effekte in der Psychologie und der Managementwissenschaft ist, dass spezifische und schwierige Ziele zu besseren Leistungen führen als vage „do your best“-Ziele.<sup>465</sup> Dieser Effekt wird mit über 90 % in verschiedener Labor- und Feldstudien nachgewiesen.<sup>466</sup> Demgegenüber stellen Ordóñez et al. (2009) jedoch fest, dass Ziele zwar die Aufmerksamkeit fokussieren, aber „[u]nfortunately, goals can focus attention so narrowly that people overlook other important features of a task“<sup>467</sup>. Die Literatur hat daher in einigen Fällen auch gezeigt,

---

<sup>464</sup> Weitere Anreize wie beispielsweise die einer ausstehenden Beförderung oder Lob und Tadel werden im Folgenden nicht weiter diskutiert.

<sup>465</sup> Vgl. Locke und Latham 2013, S. 5; Polzer und Neale 1995, S. 3.

<sup>466</sup> Vgl. Locke et al. 1981, S. 125.

<sup>467</sup> Ordóñez et al. 2009, S. 7.

dass spezifische Ziele nicht zu einer besseren Leistung führen oder die Leistung nicht beeinflussen.<sup>468</sup>

Es sind mehrere Erklärungen denkbar, warum spezifische und schwierige Ziele bisher zur Algorithmen-Aversion beigetragen haben und warum ein „do your best“-Ziel diese verringern könnte. Dies meint beispielsweise, dass durch das Setzen spezifischer, schwieriger Ziele Forecaster ihren Fokus auf vermeintlich zielrelevante Handlungen beschränken und somit weniger kognitive Ressourcen auf neue Arbeitsmethoden richten.<sup>469</sup> Dies könnte dazu führen, dass sich die Forecaster ausschließlich auf ihre eigene Leistung konzentrieren, anstatt sie sorgfältig mit der Leistung des Algorithmus zu vergleichen.

Ergänzend kritisieren Webb et al. (2013), dass anspruchsvollere Ziele die Forecaster daran hindern, neue Strategien zur Steigerung der Effizienz zu entdecken bzw. anzuwenden.<sup>470</sup> Stattdessen arbeiten sie härter mit ihren bestehenden Methoden. Das bedeutet, dass schwierige Ziele zwar die Leistung steigern, aber auch die Entdeckung neuer, effizienterer Arbeitsroutinen – wie der, einem Algorithmus den Vorzug zu geben – behindern können. Polzer und Neale (1995) führen hierzu an: „[S]pecific goals failed to appropriately incorporate new information presented during a negotiation and consequently achieved poorer outcomes than negotiators with do-your-best goal“<sup>471</sup>. Sie spezifizieren und erweitern die Sicht auf die Entdeckung neuer Methoden und Strategien dahingehend, dass Ziele Forecaster daran hindern, über bestimmte Aufgaben-dimensionen nachzudenken, die nicht direkt vom ausgegebenen Ziel betroffen sind – ergo eine sorgfältige Evaluation der Leistung des Algorithmus anstelle des Fokus auf die Verbesserung der eigenen Leistung. Hierzu ergänzen van Dyck et al. (2005), dass schwierige Ziele Forecaster daran hindern, neue Informationen innerhalb einer Aufgabe effektiv zu verarbeiten.<sup>472</sup> Die Leistung des Algorithmus unterliegt einer gewissen Schwankung, sodass mit jeder Übungsrunde eine neue Information verarbeitet und eingeordnet werden muss.

Latham und Locke (2013) stellen außerdem fest, dass die Angst, Fehler zu machen, aufgrund schwieriger Ziele zunimmt. Sie schlagen vor, dass Fehler akzeptiert

---

<sup>468</sup> Vgl. Akın und Karagözoğlu 2017, S. 193; Seijts et al. 2013, S. 207.

<sup>469</sup> Siehe dazu auch Earley et al. 1989; Wood et al. 1990.

<sup>470</sup> Vgl. Webb et al. 2013, S. 1433.

<sup>471</sup> Polzer und Neale 1995, S. 7.

<sup>472</sup> Vgl. van Dyck et al. 2005, S. 1228.

und gefördert werden müssen, damit die Forecaster den Fehler richtig einschätzen und daraus lernen können. Eine Arbeitsweise, in der Fehler akzeptiert und zeitweise sogar gefördert werden, wird als Fehlerkultur bezeichnet.<sup>473</sup> Fehlerkultur kann positiv mit Leistung und Zielerreichung verbunden sein.<sup>474</sup> Ein solches Arbeitsumfeld, in dem Fehler akzeptiert werden, kann mit einem „do your best“-Ziel geschaffen werden. Ein „do your best“-Ziel, bei dem den Forecastern mitgeteilt wird, dass sie ihr Incentive unabhängig von ihrer Leistung erhalten, soll sich demnach positiv auf den Einsatz von Algorithmen auswirken.<sup>475</sup> Es wird daher argumentiert, dass schwierige Ziele die Forecaster daran hindern, die überlegene, aber nicht perfekte Leistung des Algorithmus hinreichend zu verarbeiten, da sich die Forecaster nur auf die Fehler des Algorithmus und die Entwicklung der eigenen Leistung konzentrieren. Die Einführung eines „do your best“-Ziels hingegen, kann den Einsatz von Algorithmen fördern und damit die Algorithmen-Aversion abschwächen:

H3: Ein „do your best“-Ziel reduziert die Algorithmen-Aversion.

#### **4.2.4 Der Einfluss von Entscheidungsrechten für Eingabeinformationen auf die Algorithmen-Aversion**

Im Folgenden wird gezeigt, welcher Einfluss von Entscheidungsrechten in der Informationsverarbeitung eines Algorithmus auf die Erstellung von Forecasts zu erwarten ist und welche Auswirkungen sich auf die Algorithmen-Aversion absehen lassen.

Es ist bedeutsam zu hinterfragen, wie Forecaster Informationen für ihren eigenen Forecast und den des Algorithmus handhaben und bewerten.<sup>476</sup> In der Literatur zur Algorithmen-Aversion wird ein spezifischer, meist nur kurz beschriebener Algorithmus verwendet.<sup>477</sup> Den Forecastern in den Experimenten wird außerdem die gemeinsame Datenbasis beschrieben, mit der sowohl sie selbst als auch der Algorithmus arbeiten.<sup>478</sup> Durch die Interaktion mit der Datenbank in den Übungsforecasts bekommen die Forecaster einen Eindruck davon, wie sie aus den Daten einen effektiven Forecast ableiten können – sie lernen dadurch, ihren eigenen Forecast zu verbessern. Dabei

---

<sup>473</sup> Vgl. Gold et al. 2014, S. 303; Seckler et al. 2017, S. 21.

<sup>474</sup> Vgl. van Dyck et al. 2005, S. 1228.

<sup>475</sup> Vgl. Gold et al. 2014, S. 303; van Dyck et al. 2005, S. 1231.

<sup>476</sup> Vgl. Bonner und Sprinkle 2002, S. 303.

<sup>477</sup> Siehe Abschnitt 2.1.1.

<sup>478</sup> Siehe Abschnitt 3.2.1.

wird den aufmerksamen Forecastern ersichtlich, dass nicht alle Daten die gleiche Relevanz für die Genauigkeit eines Forecasts haben. Sie lernen demnach während der Übungsforecasts, welche Daten für einen guten Forecast besonders relevant sind. Mit diesem Wissen über die Relevanz von Input-Daten beurteilen sie auch die Leistung und womöglich die ungenaueren Forecasts des Algorithmus. Dieser kognitive Prozess ist im Forecasting umfassend untersucht: „It is a widespread forecasting practice to judgmentally adjust statistical forecasts in order to improve forecast accuracy“<sup>479</sup>. Burton et al. (2019) konstatieren daher für die Interaktion mit Algorithmen: „For a human decision maker to act on an algorithm's judgement, he or she must feel in control and confident enough to place trust in it“<sup>480</sup>. Dies impliziert, dass es einen gewissen Bedarf für Forecaster bei der Auswahl der Informationen und Daten gibt, wenn Forecaster mit Algorithmen interagieren.

In der Literatur finden sich Hinweise darauf, dass eine zunehmende Interaktion zwischen Forecaster und Algorithmus zu einer sinkenden Algorithmen-Aversion führen kann. Fildes et al. (2009) zeigen, dass 90 % der Ergebnisse von Algorithmen in Forecasts nicht direkt akzeptiert werden, sondern vom Forecaster verändert werden. Dies zeigt den Wunsch der Forecaster nach Entscheidungsrechten bei der Interaktion mit Algorithmen. Damit im Einklang zeigen Dietvorst et al. (2018), dass die Algorithmen-Aversion gesenkt werden kann, indem Forecaster die von einem Algorithmus erzeugten Forecasts modifizieren dürfen. Der Fokus dieser Modifikationen liegt zumeist in der Anpassung des Ergebnisses des Algorithmus. Die Änderungen, die ein Forecaster am Ergebnis eines Algorithmus vornimmt, unterliegt jedoch den klassischen Biases<sup>481</sup>, wodurch ein ungewollter Effekt eintritt: Die Anpassung des Ergebnisses des Algorithmus führt dazu, dass die Genauigkeit der Vorhersage abnimmt.<sup>482</sup> Eroglu und Croxton (2010) mahnen diesbezüglich an, dass „[...] forecaster may see a pattern in data even though none exists“<sup>483</sup>. Fildes et al. (2009) sehen, dass kleine Änderungen an Forecasts deren Genauigkeit verschlechtern, während größere Änderungen oft zu einer Verbesserung führen.<sup>484</sup> Insbesondere in Fällen, in denen Forecaster glauben, dass

---

<sup>479</sup> Eroglu und Croxton 2010, S. 116.

<sup>480</sup> Burton et al. 2019, S. 4.

<sup>481</sup> Siehe Abschnitt 2.1.4.1 und Lawrence et al. 2006, S. 495–498.

<sup>482</sup> Vgl. Carbone et al. 1983, S. 559; Dietvorst et al. 2018, S. 1169; Goodwin und Fildes 1999, S. 37; Remus et al. 1995, S. 290.

<sup>483</sup> Eroglu und Croxton 2010, S. 117.

<sup>484</sup> Hier sei auf externe Schocks wie sie zum Beispiel durch die Coronapandemie entstanden sind verwiesen. Siehe dazu auch Abschnitt 6.2.2.

sie in den Daten negative Effekte sehen, die der Algorithmus nicht berücksichtigen kann, kann es sinnvoll sein, dass Forecaster diese Informationen vor der Informationsverarbeitung aus der Datenbasis entfernen.<sup>485</sup>

Wenn ein Forecaster Erfahrungen mit den für eine Prognose zugrundeliegenden Daten sammelt, könnte er den Eindruck gewinnen, dass der Algorithmus die Informationen falsch gewichtet. Es wird demnach der Drang entstehen, das Ergebnis des Algorithmus anzupassen. Die Anpassung des Ergebnisses wird jedoch in aller Regel den Forecast verschlechtern.<sup>486</sup> Es wird demzufolge die Hypothese aufgestellt, dass die Algorithmen-Aversion ebenfalls abgeschwächt wird, wenn ein Forecaster die Informationen, die der Algorithmus verarbeitet, ex ante beeinflussen kann:

H4: Entscheidungsrechte für Eingabeinformationen, die der Algorithmus verarbeitet, reduzieren die Algorithmen-Aversion.

#### 4.2.5 Zusammenfassung der Hypothesen

Tabelle 4 fasst die erwarteten Effekte in den Hypothesen von Studie 1 nochmals zusammen. Während H1 die Algorithmen-Aversion in dieser Arbeit nochmals bestätigen soll, zielen H2–H4 auf die Reduktion der Algorithmen-Aversion durch das Arbeitsumfeld ab. Im Folgenden wird beschrieben, welche notwendigen Änderungen im Experimentaldesign zur Untersuchung dieser Hypothesen vorgenommen werden müssen und wie die Hypothesen operationalisiert werden.

Tabelle 4: Zusammenfassung der Hypothesen in Studie 1

Hypothese	Prognostizierte Wirkung
H1: Durch Leistungsfeedback entsteht eine Algorithmen-Aversion.	↑
H2: Zeitdruck reduziert die Algorithmen-Aversion.	↓
H3: Ein „do your best“-Ziel reduziert die Algorithmen-Aversion.	↓
H4: Entscheidungsrechte für Eingabeinformationen, die der Algorithmus verarbeitet, reduzieren die Algorithmen-Aversion.	↓

<sup>485</sup> Vgl. Remus et al. 1995, S. 285.

<sup>486</sup> Vgl. Dietvorst et al. 2018, S. 1155; Fildes et al. 2009, S. 3.

### 4.3 Experiment und Beschreibung der Datengrundlage

#### 4.3.1 Experimentaldesign und Operationalisierung der Manipulation

Entsprechend den Ausführungen in Kapitel 3 wurde zunächst das Experimentaldesign von Dietvorst et al. (2015) repliziert und anschließend erweitert.<sup>487</sup> Zum Ablauf des Experiments, den relevanten Aspekten wie der Forecastingaufgabe und der Incentivierung siehe Abschnitt 3.2. Im Folgenden wird mit Fokus auf die eingeführten Treatments und die damit verbundenen Manipulationen beschrieben, wie das Experiment in Studie 1 abgelaufen ist. Dabei wird zunächst erklärt, wie die Zuteilung der Teilnehmer zu den unterschiedlichen Experimentalgruppen erfolgt ist und wie die Treatments operationalisiert wurden. Schließlich war es aufgrund des Replikationscharakters der Forecastingaufgabe und der Teilnehmer (MTurker) notwendig, einige Erweiterungen am Experiment vorzunehmen. Diese sind in Abschnitt 3.2 erläutert.

Im Experimentalsetting werden die Teilnehmer nach dem Zufallsprinzip einer von acht Gruppen zugewiesen. In Tabelle 5 werden diese acht Gruppen in eine eindeutige Nomenklatur überführt.

Tabelle 5: Nomenklatur der Experimentalgruppen in Studie 1

Bezeichnung	Leistungsfeedback	Treatment
<b>APC</b>	Algorithmen-Appreciation (AP)	Kontroll (C)
<b>APTP</b>	Algorithmen-Appreciation (AP)	Zeitdruck (TP)
<b>APDB</b>	Algorithmen-Appreciation (AP)	„do your best“-Ziel (DB)
<b>APDR</b>	Algorithmen-Appreciation (AP)	Decision Rights (DR)
<b>AAC</b>	Algorithmen-Aversion (AA)	Kontroll (C)
<b>AATP</b>	Algorithmen-Aversion (AA)	Zeitdruck (TP)
<b>AADB</b>	Algorithmen-Aversion (AA)	„do your best“-Ziel (DB)
<b>AADR</b>	Algorithmen-Aversion (AA)	Decision Rights (DR)

Zum besseren Verständnis von Tabelle 5 kann die schematische Darstellung in Abbildung 31 herangezogen werden. Die acht Gruppen lassen sich entsprechend der Grundlagen zur Algorithmen-Aversion entlang zweier wesentlicher Pfade (blaue Kästen) unterteilt in zwei Vierergruppen (gelber Kasten) subsumieren. Ein Pfad sam-

<sup>487</sup> Siehe Abschnitt 4.3.

## Studie 1

melte keine Erfahrungen mit dem Algorithmus und der eigenen Forecastinggenauigkeit – ergo liegt hier kein Leistungsfeedback vor (unterer blauer Kasten). Entsprechend der Literatur wird den vier Gruppen, die diesem Pfad folgen, der Terminus Algorithmen-Appreciation (AP) vorangestellt. Die anderen vier Gruppen sammelten Erfahrungen mit dem Algorithmus und dem eigenen Forecast – ergo erhalten sie Leistungsfeedback (oberer blauer Kasten). Ihnen wird in der Nomenklatur entsprechend die Algorithmen-Aversion (AA) vorangestellt. In jeweils einer Gruppe je Pfad wurden keine weiteren Anpassungen vorgenommen. Diese Gruppen entsprechen Dietvorst et al. (2015) und tragen zusätzlich den Buchstaben C für *Kontrollgruppe*.<sup>488</sup> Der Vergleich dieser beiden Gruppen (APC und AAC) ermöglicht die Untersuchung von H1.

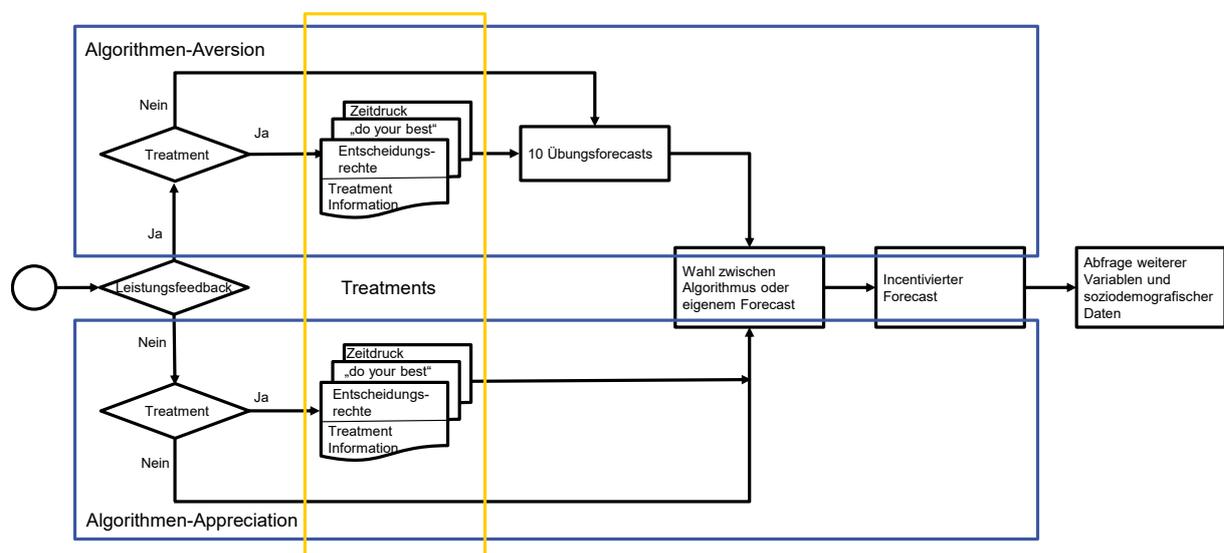


Abbildung 31: Schematische Darstellung des Ablaufs des Experiments<sup>489</sup>

Das Treatment bzw. die Operationalisierung von Zeitdruck ist die Begrenzung der verfügbaren Zeit für die Erstellung eines Forecasts. Das heißt, die Teilnehmer hatten 12 Sekunden Zeit, um eine Entscheidung zu treffen – sowohl in den Übungsforecasts als auch im incentivierten Forecast. Für die Teilnehmer im Algorithmen-Appreciation-Pfad gab es demnach keine Möglichkeit, den Zeitdruck zu erfahren und zu wissen, welche Schwierigkeit für die Erstellung des Forecasts daraus erwächst.

Da es für die vorgesehene Forecastingaufgabe keine Referenz zum tatsächlichen Zeitbedarf gab, wurde ein Pretest durchgeführt. Dabei wurden drei Aspekte berücksichtigt. Einerseits musste der Zeitdruck so niedrig sein, dass die Teilnehmer auch

<sup>488</sup> Zu den Abweichungen vom Aufbau nach Dietvorst et al. 2015 siehe Abschnitt 3.2.

<sup>489</sup> Eigene Darstellung.

## Studie 1

unter Zeitdruck noch in der Lage waren, die Aufgabe zu erfüllen. Andererseits musste der Zeitdruck groß genug sein, damit er die Bedenkzeit in der Aufgabe verkürzt – also als Zeitdruck wahrgenommen wird. Ersteres ist insbesondere für die Teilnehmer ohne Leistungsfeedback bzw. Übungsforecasts relevant, da sie keine Möglichkeit haben, sich vor dem incentivierten Forecast an den Zeitdruck zu gewöhnen. Schließlich sollte zwischen den Algorithmen-Aversion-Gruppen und den Algorithmen-Appreciation-Gruppen keine Verzerrung durch die Vorgabe von unterschiedlich großem Zeitdruck entstehen. Die beiden Gruppen, die dieses Treatment erhalten haben, werden als *Time Pressure*, kurz *TP* bezeichnet

Dazu wurde das Experiment unter 198 Studenten auf der Plattform *SurveyCircle* durchgeführt.<sup>490</sup> Als Incentive in diesem Pretest wurden zehn Amazon-Gutscheine über jeweils 5 € eingesetzt. Diese wurden den Teilnehmern mit den besten Forecasts in Aussicht gestellt. Es wurde das Originalexperiment mit Zeitdruck in drei unterschiedlichen Ausprägungen durchgeführt. Den Teilnehmern wurde dabei einmalig eingeblendet, dass sie unter zwölf, zehn oder acht Sekunden Zeitdruck stehen werden. Um sicherzustellen, dass der Zeitdruck auch als solcher wahrgenommen wird, wurden die Teilnehmer gebeten, ihre Einschätzung zum wahrgenommenen Zeitdruck zu geben.<sup>491</sup> Tabelle 6 zeigt, dass es in den Gruppen mit Übungsforecasts für alle drei Formen des Zeitdrucks zu keinen nennenswerten Ausfällen kommt. Das heißt, dass die Teilnehmer in der Lage sind, den incentivierten Forecast durchzuführen. Für die Teilnehmer ohne Übungsforecasts zeigt sich hingegen, dass der Zeitdruck von acht und zehn Sekunden eine zu große Herausforderung ist und nahezu 30 % der Teilnehmer den incentivierten Forecast nicht schaffen. Dies spricht für einen Zeitdruck von zwölf Sekunden (13 % Ausfälle).

Tabelle 6: Ausfälle durch Zeitdruck im Pretest

<b>Zeitdruck in Sekunden,</b>	<b>12</b>	<b>10</b>	<b>8</b>
<b>Algorithmen- Aversion</b>	1 aus 31 (3 %)	0 aus 32 (0 %)	0 aus 33 (0 %)
<b>Algorithmen- Appreciation</b>	4 aus 32 (13 %)	10 aus 34 (29 %)	10 aus 36 (28 %)

<sup>490</sup> Zu erreichen unter: <https://www.surveycircle.com/de/> (zuletzt geprüft am 31.05.2022).

<sup>491</sup> Auf eine 5er Likert-Skala mit den Merkmalsausprägungen: Very Low, Low, Medium, High, Very High.

## Studie 1

Tabelle 7 und Abbildung 32 zeigen, wie der wahrgenommene Zeitdruck von acht auf zwölf Sekunden deutlich abnimmt, jedoch auf einem hohen Niveau von 3,78 bzw. 3,97 verbleibt.<sup>492</sup> Es wurde daher ein Zeitdruck von zwölf Sekunden festgelegt, da dieser als Zeitdruck wahrgenommen wird und dennoch genug Zeit bietet, um die Aufgabe erfüllen zu können.

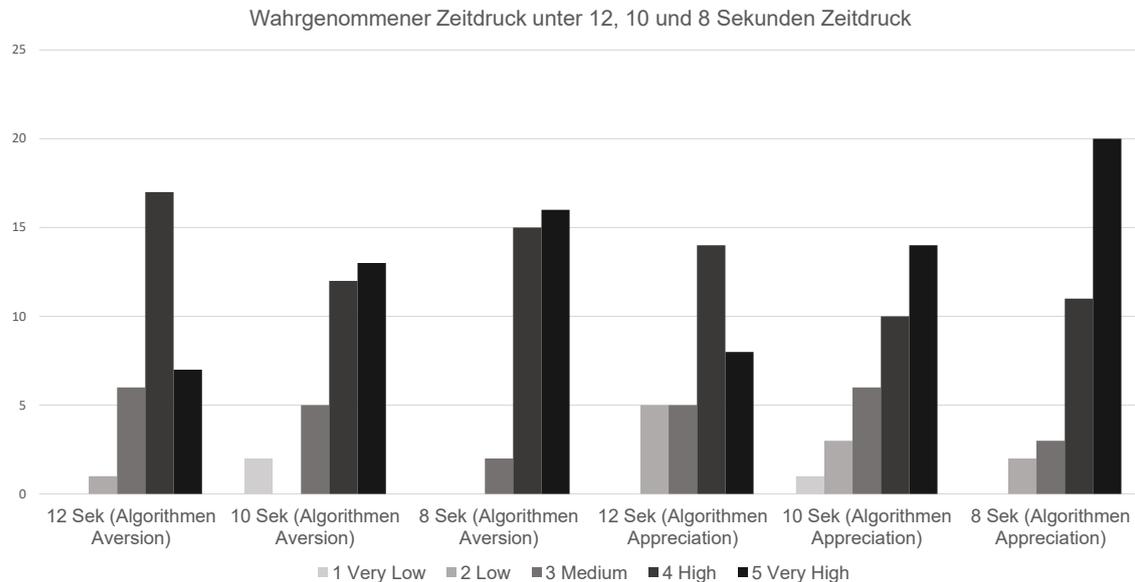


Abbildung 32: Wahrnehmung von unterschiedlich starkem Zeitdruck<sup>493</sup>

Tabelle 7: Mittelwerte des wahrgenommenen Zeitdrucks (M und SD)

Zeitdruck in Sekunden	12	10	8
Algorithmen-Aversion	3,97 (0,75)	4,06 (1,08)	4,42 (0,61)
Algorithmen-Appreciation	3,78 (1,00)	3,97 (1,11)	4,36 (0,87)

<sup>492</sup> Auf einer 5er Likert-Skala, siehe Abbildung 32.

<sup>493</sup> Eigene Darstellung.

## Studie 1

Damit sich die Teilnehmer auf den Zeitdruck im incentivierten Forecast einstellen können, wurde dieser im Experiment kenntlich gemacht (siehe Abbildung 33).

Now, you will go through 1 official estimate.

For your estimate you will receive the data shown as below. As you only have a limited amount of time available for your estimation, please read through this very carefully and make sure you understand each category and get comfortable with the way the data is displayed.

The data will be shown as follows:

**Remaining Time: XX**

Number of Major Airports	XX
Census Population Rank - 2010	XX
Number of Counties Rank	XX
Median Household Income Rank - 2008	XX
Domestic Travel Expenditure Rank - 2009	XX

What do you think this state's rank is in terms of departing airline passengers in 2011? (Please enter a number 1-50)

By clicking on the next button your official estimation starts with 12 seconds time limit.

Abbildung 33: Zeitdruck in der Algorithmen-Appreciation<sup>494</sup>

Den Teilnehmern in der Bedingung mit dem „do your best“-Ziel wird mitgeteilt, dass sie ein fixes Incentive in Höhe von 2 \$ erhalten werden, unabhängig von der Genauigkeit ihres incentivierten Forecasts. Sie wurden aufgefordert ihr Bestes zu geben, um bei der anstehenden Forecastingaufgabe den bestmöglichen Forecast zu erzielen (siehe Abbildung 34). Die beiden Gruppen, die dieses Treatment erhalten haben, werden als „do your best“, kurz *DB* bezeichnet. Entsprechend wurde für diese Teilnehmer jeder weitere Bezug zu einer leistungsabhängigen Vergütung innerhalb des Experiments ausgeblendet (siehe Anhang 6)

We want you to **do your best in this task** to get the best estimation.

Therefore you will go through 10 practice rounds to get experience with the data.

Abbildung 34: Darstellung des „do your best“-Ziels<sup>495</sup>

Den Teilnehmern der Bedingung mit Entscheidungsrechten für Eingabeinformationen wird die Bezeichnung *Decision Rights* bzw. *DR* vorangestellt. Den Probanden

<sup>494</sup> Eigene Darstellung (Auszug aus dem Experiment).

<sup>495</sup> Eigene Darstellung (Auszug aus dem Experiment).

## Studie 1

in dieser Gruppe wurde mitgeteilt, dass der Algorithmus vier der fünf gezeigten Variablen verarbeiten könne und dass sie diese Auswahl treffen müssen. Die Incentivierung war die gleiche wie bei den Kontrollgruppen. Die Information zur Auswahl der zu verarbeitenden Daten wurde im Pfad der Algorithmen-Aversion bereits vor den zehn Übungsforecasts gegeben. Das heißt, vor den Trainingsrunden haben die Probanden eine Auswahl von vier aus fünf zu verarbeitenden Daten durchgeführt. Dadurch bestand die Möglichkeit, dass ein Teilnehmer, nachdem er sich zunächst für vier Datenpunkte entschieden hatte, diese Entscheidung nach den ersten zehn Runden nochmals überdenken konnte. Da so ein Risiko einer Aversion gegenüber dem Algorithmus durch eine vermeintlich falsche Datenauswahl entstehen kann, wurde den Teilnehmern die Möglichkeit eingeräumt, ihre ursprüngliche Auswahl vor dem incentivierten Forecast nochmals anzupassen (siehe Abbildung 35).<sup>496</sup> Für die Hypothesen H2–H4 werden die Unterschiede der drei Treatmentbedingungen innerhalb der Algorithmen-Aversion, also mit der AAC, verglichen.

You can make changes to you prior choice for the models 4 input factors.  
Do you wish to make changes to the 4 input factors for the model?

No  Yes

Abbildung 35: Möglichkeit der Anpassung der der Eingabeinformationen<sup>497</sup>

### 4.3.2 Deskriptive Statistik

Die Auswahl, Bezahlung und Gruppengröße, folgt im Wesentlichen dem Vorgehen in Dietvorst et al. (2015). Die Erhebung wurde am 02.03.2020 durchgeführt. Es werden in Summe 3032 Teilnehmer akquiriert. Aus dieser Gesamtheit werden 865 Teilnehmer ausgeschlossen, da sie die wesentliche abhängige Variable zur Entscheidung zwischen dem eigenen Forecast oder dem des Algorithmus nicht beantwortet haben. 47 Teilnehmer werden zudem ausgeschlossen, weil sie die Aufmerksamkeitsprüfung, und weitere 280 Teilnehmer, weil sie die Verständnisfragen nicht hinreichend beantwortet haben (siehe dazu Abschnitt 3.2). Somit verbleibt eine Stichprobe von 1840 Teilnehmern.<sup>498</sup> In dieser Stichprobe liegt ein ausgeglichenes Verhältnis von Frauen und Männern mit 46 % weiblichen und 54 % männlichen Teilnehmern vor. Die Stichprobe weist ein Durchschnittsalter von 39 Jahren auf (SD = 11,51) und erstreckt sich über eine

<sup>496</sup> Weitere Analysen hierzu erfolgen in Abschnitt 4.4.3.3.

<sup>497</sup> Eigene Darstellung (Auszug aus dem Experiment).

<sup>498</sup> Dies entspricht eine Ausschlussrate von 40 %.

## Studie 1

Altersspanne von 18 bis 76 Jahren (siehe Abbildung 36). Dabei sind über 50 % der Teilnehmer unter 35 Jahre alt.

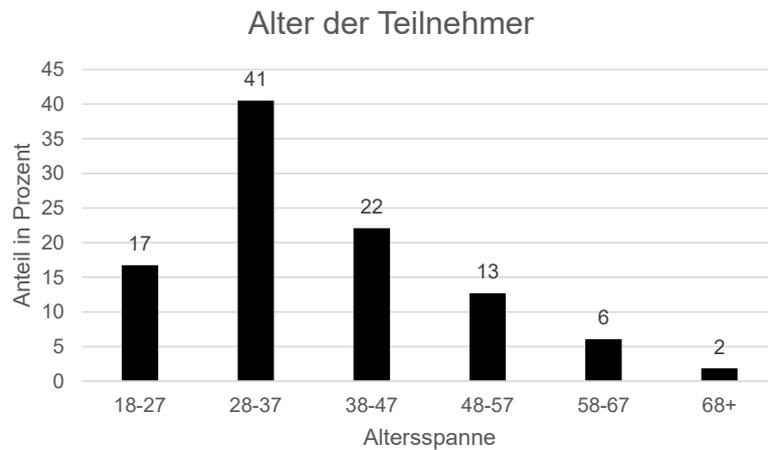


Abbildung 36: Alter der Teilnehmer in Studie 1<sup>499</sup>

Abbildung 37 zeigt die Verteilung des jeweils höchsten Bildungsabschlusses der Teilnehmer in der finalen Stichprobe (ansteigend v. l. n. r.). Es ist ersichtlich, dass der Großteil der Teilnehmer mindestens ein vierjähriges College besucht hat. Das vierjährige College ist einem europäischen Vordiplom oder Bachelorabschluss gleichzusetzen.<sup>500</sup>

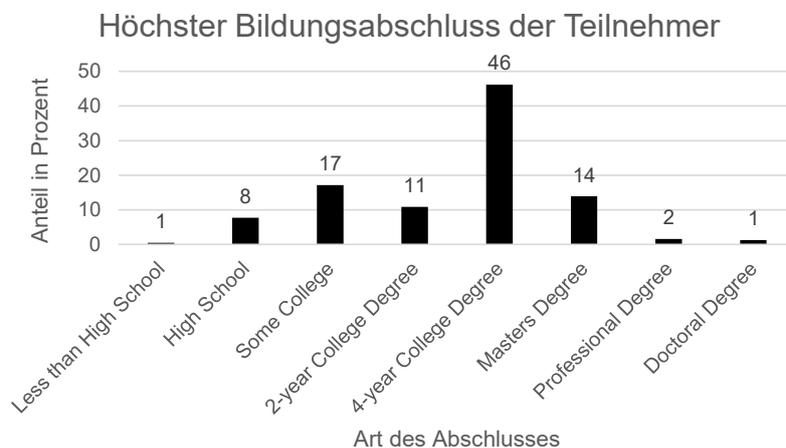


Abbildung 37: Bildungsniveau der Teilnehmer in Studie 1<sup>501</sup>

Zu besseren Einordnung und Vergleichbarkeit der Stichprobe werden die deskriptiven Werte in Tabelle 8 Dietvorst et al. (2015) gegenübergestellt. Es zeigt sich,

<sup>499</sup> Eigene Darstellung.

<sup>500</sup> Vgl. Büchtemann et al. 1993, S. 510.

<sup>501</sup> Eigene Darstellung.

## Studie 1

dass die Stichprobe in dieser Arbeit im Schnitt fünf Jahre älter ist. Die Verteilung des Geschlechts bewegt sich in einem ähnlichen Bereich. Ebenso wird eine ähnlich große Gruppengröße in den Experimentalgruppen gezeigt.

Tabelle 8: Vergleich der Stichprobe aus Studie 1 mit Dietvorst et al. (2015)<sup>502</sup>

	<b>Studie 1</b>	<b>Dietvorst et al. (2015)</b>
<b>Durchschnittsalter</b>	37 – 40	33 – 34
<b>Geschlecht (% weiblich)</b>	43 – 49	46 – 53
<b>Stichprobenumfang (durchschnittliche Gruppengröße)</b>	230	259

### 4.3.2.1 Zuteilung der Experimentalgruppen

Die zufällige Verteilung der Stichprobe in die Experimentalgruppen ergibt die finalen Gruppengrößen in Tabelle 9. Die Teilnehmer wurden randomisiert den Gruppen zugewiesen. Die Regel für die Randomisierung lautete *gleichverteilte Ziehung*. Zu Beginn der Erhebung starteten sehr viele Teilnehmer fast zeitgleich das Experiment. Das heißt, wenn in einer Gruppe die Teilnehmer zuverlässiger und/oder schneller waren, wurden die Gruppen hier relativ gesehen etwas größer.

Tabelle 9: Gruppeneinteilung in Studie 1

	<b>Kontroll- gruppe (C)</b>	<b>Zeitdruck (TP)</b>	<b>„do your best“ (DB)</b>	<b>Entscheidungs- rechte (DR)</b>
<b>Algorithmen-Appreciation (AP)</b>	n = 219	n = 263	n = 208	n = 209
<b>Algorithmen-Aversion (AA)</b>	n = 219	n = 223	n = 237	n = 262

Tabelle 10 zeigt die soziodemografischen Daten zu Geschlecht und Alter in Studie 1 aufgeteilt nach Experimentalgruppen. Tabelle 11 ergänzt die die soziodemografischen Daten zwischen den Experimentalgruppen um den Bildungsabschluss.

<sup>502</sup> Eine Spanne entsteht durch die Angabe über mehrere Versuchsgruppen hinweg. Siehe Dietvorst et al. 2015, S. 116.

## Studie 1

Tabelle 10: Geschlecht und Alter nach Experimentalgruppen in Studie 1

Gruppe	Geschlecht (% weiblich)	Durchschnittsalter
APC (n = 219)	49	40,21
APTP (n = 263)	47	39,95
APDB (n = 208)	43	39,33
APDR (n = 209)	44	38,26
AAC (n = 219)	47	40,51
AATP (n = 223)	44	38,48
AADB (n = 237)	45	37,91
AADR (n = 262)	45	38,65

Tabelle 11: Bildungsabschluss nach Experimentalgruppen in Studie 1<sup>503</sup>

Gruppe	Bildungsabschluss (Anteil in %)
APC (n = 219)	1 = 0,5; 2 = 7,8; 3 = 18,4; 4 = 12,4; 5 = 46,1; 6 = 11,1; 7 = 2,3; 8 = 1,4
APTP (n = 263)	1 = 0; 2 = 7,7; 3 = 18,9; 4 = 11,2; 5 = 44,4; 6 = 16,2; 7 = 0,4; 8 = 1,2
APDB (n = 208)	1 = 0; 2 = 8,7; 3 = 14,6; 4 = 8,7; 5 = 49,0; 6 = 16,5; 7 = 1,5; 8 = 1,0
APDR (n = 209)	1 = 0,5; 2 = 6,8; 3 = 15; 4 = 11,7; 5 = 48,5; 6 = 15,5; 7 = 1,5; 8 = 0,5
AAC (n = 219)	1 = 0,5; 2 = 8,2; 3 = 20,1; 4 = 9,1; 5 = 43,8; 6 = 14,6; 7 = 3,2; 8 = 0,5
AATP (n = 223)	1 = 0,9; 2 = 5,8; 3 = 20,2; 4 = 10,8; 5 = 44,8; 6 = 13,9; 7 = 1,3; 8 = 2,2
AADB (n = 237)	1 = 0,8; 2 = 9,3; 3 = 14,0; 4 = 12,3; 5 = 48,3; 6 = 11,4; 7 = 2,1; 8 = 1,7
AADR (n = 262)	1 = 0,8; 2 = 7,6; 3 = 17,2; 4 = 11,5; 5 = 47,3; 6 = 13,0; 7 = 1,1; 8 = 1,5

### 4.3.2.2 Performance und Bezahlung

Damit eine Algorithmen-Aversion auch als solche gesehen werden kann, müssen die Teilnehmer in ihrer Forecastinggenauigkeit schlechter abschneiden als der Algorithmus. Andererseits würden sie durch ihren eigenen Forecast einen höheren Bonus realisieren und es wäre rational, dass sie den Algorithmus ablehnen. Tabelle 12 zeigt, dass es für die Teilnehmer in allen Gruppen sinnvoll gewesen ist, den Algorithmus zu

<sup>503</sup> 1 = Less than High School; 2 = High School; 3 = Some College; 4 = 2-year College Degree; 5 = 4-year College Degree; 6 = Masters Degree; 7 = Professional Degree; 8 = Doctoral Degree.

## Studie 1

wählen und dass eine Untersuchung der Algorithmen-Aversion fortgesetzt werden kann:

*Tabelle 12: Forecastinggenauigkeit von Forecastern und Algorithmus in Studie 1 (M und SD)*

<b>Gruppe</b>	<b>Genauigkeit Forecaster</b>	<b>Genauigkeit Algorithmus</b>	<b>Differenz</b>	<b>Paardifferenzentest</b>
<b>APC</b>	11,43 (9,52)	4,44 (4,35)	6,99 (10,01)	t(218) = 10,332 p < 0,001
<b>APTP</b>	13,79 (11,49)	4,23 (4,11)	9,56 (11,80)	t(260) = 13,083 p < 0,001
<b>APDB</b>	10,95 (10,76)	3,69 (3,55)	7,26 (11,33)	t(207) = 9,240 p < 0,001
<b>APDR</b>	10,33 (9,03)	4,72 (4,20)	5,61 (9,69)	t(208) = 8,360 p < 0,001
<b>AAC</b>	8,78 (8,72)	4,31 (4,36)	4,47 (8,79)	t(218) = 7,519 p < 0,001
<b>AATP</b>	9,05 (9,23)	4,56 (4,36)	4,49 (9,91)	t(222) = 6,772 p < 0,001
<b>AADB</b>	7,79 (6,81)	4,41 (4,09)	3,39 (7,40)	t(236) = 7,048 p < 0,001
<b>AADR</b>	7,81 (7,51)	4,39 (3,97)	3,43 (8,22)	t(261) = 6,747 p < 0,001

Tabelle 12 zeigt, dass es über alle Gruppen hinweg einen signifikanten Unterschied in der Forecastinggenauigkeit zwischen den Forecastern und dem Algorithmus gibt. Während die Forecasts des Algorithmus im Mittel zwischen 3,69 und 4,72 schwanken, liegen die der Forecaster zwischen 7,79 und 13,79. Auffällig sind dabei die Unterschiede zwischen den Algorithmen-Appreciation- und den Algorithmen-Aversion-Gruppen.<sup>504</sup> Hier zeigt sich der Lerneffekt, der durch die zehn Übungsrunden mit Leistungsfeedback erreicht wird. Besonders schlecht schneiden die Teilnehmer unter Zeitdruck und ohne Übungsforecasts ab. Jeweils besser als die Kontrollgruppen sind die Teilnehmer in der DB- und DR-Gruppe, wobei die absolut beste Performance in der AADB-Gruppe mit Leistungsfeedback vorliegt.

<sup>504</sup> Ein t-Test zur Genauigkeit der Forecaster zwischen Appreciation-Gruppe und Aversion-Gruppe zeigt hier, dass sich die Leistung der Teilnehmer durch die Übungsrunden signifikant verbessert: Algorithmen-Appreciation (M und SD): 11,75 (10,39). Algorithmen-Aversion (M und SD): 8,33 (8,092). t(1836) = 7,896, p < 0,001.

## Studie 1

Analog zu Tabelle 12 findet sich in Tabelle 13 der jeweils aus der Leistung resultierende Bonus. Es wird abermals der Vergleich zur durch den Algorithmus möglichen Vergütung aufgezeigt.

Tabelle 13: Bonuszahlungen in \$ in Studie 1 (M und SD)<sup>505</sup>

Gruppe	Bonus Forecaster	Bonus Algorithmus	Algo-Differenz	Paardifferenzentest
<b>APC</b>	0,23 (0,32)	0,49(0,37)	-0,26 (0,47)	t(218) = -8,126 p < 0,001
<b>APTP</b>	0,22 (0,32)	0,49(0,37)	-0,28 (0,44)	t(260) = -10,166 p < 0,001
<b>APDB</b>	0,29(0,36) <sup>506</sup>	0,53(0,34)	-0,24 (0,49)	t(207) = -7,012 p < 0,001
<b>APDR</b>	0,25(0,34)	0,45 (0,36)	-0,19 (0,45)	t(208) = -6,294 p < 0,001
<b>AAC</b>	0,31 (0,35)	0,50 (0,36)	-0,19 (0,42)	t(218) = -6,737 p < 0,001
<b>AATP</b>	0,30 (0,34)	0,47(0,37)	-0,17 (0,46)	t(222) = -5,506 p < 0,001
<b>AADB</b>	0,31 (0,35) <sup>507</sup>	0,48(0,36)	-0,17 (0,45)	t(236) = -5,840 p < 0,001
<b>AADR</b>	0,33 (0,36)	0,47(0,37)	-0,14 (0,48)	t(261) = -4,722 p < 0,001

Die nicht direkt übertragbare Leistung von der Genauigkeit des Forecasts in Tabelle 12 und der schlussendlichen Bezahlung in Tabelle 13, begründet sich in der Verteilung der Performance in Verbindung mit der Regel zur Incentivierung (siehe Abschnitt 3.2.2 und 3.2.3). Dadurch, dass nur bis zu einer Genauigkeit bzw. Abweichung von sechs Rangpositionen eine leistungsabhängige Vergütung gegeben wurde, verlaufen Bezahlung und Performance nicht synchron.<sup>508</sup> Für APDB sind beispielsweise 51 % der Teilnehmer nicht in den Bereich hervorgehoben, für den es eine leistungsabhängige Vergütung gab. Für die APDR liegt dieser Wert bei 59 %. Darin begründet

<sup>505</sup> Bei der Interpretation von Tabelle 13 ist zu berücksichtigen, dass die Teilnehmer, die außerhalb des Bereichs der Bonuszahlungen lagen, dennoch eine Bezahlung i. H. v. 0,01 \$ erhalten haben. Das heißt, alle Forecasts außerhalb des Bonusbereichs gehen sowohl für den Algorithmus, als auch den menschlichen Forecast, mit 0,01 \$ in die Berechnung ein.

<sup>506</sup> De facto bedingt das „do your best“-Ziel einen 1 \$ Bonus. Den Teilnehmern wurde entsprechend 1 \$ ausbezahlt.

<sup>507</sup> De facto wurde 1 \$ Bonus verdient.

<sup>508</sup> Korrelation nach Pearson zeigt zwischen Leistung Forecaster und Vergütung Forecaster:  $r = -0,679$ ,  $p < 0,001$  und zwischen Leistung Algorithmus und Vergütung Algorithmus:  $r = -0,879$ ,  $p < 0,001$ .

## Studie 1

sich die Differenz in der Vergütung trotz ähnlicher mittlerer Leistung.<sup>509</sup> Dieser Zusammenhang wird in Anhang 3 und Anhang 4 nochmals verdeutlicht. Die Differenzen in der Vergütung innerhalb der AA- und der AP-Gruppe sind darüber hinaus nicht signifikant und werden daher nicht weiter betrachtet.<sup>510</sup>

### 4.3.2.3 Algorithmen-Aversion und Confidence

Aufgrund des bedeutsamen Effekts der Confidence, wie in Abschnitt 2.1.4.3 beschrieben, werden die Werte zur Confidence in den eigenen Forecast und die in den Forecast des Algorithmus für jede Experimentalgruppe in Tabelle 14 beschrieben.

Tabelle 14: Confidencewerte in Studie 1 (M und SD)

Gruppe	Confidence Forecaster	Confidence Algorithmus	Differenz	Paardifferenztest
<b>APC</b>	2,99 (0,98)	3,79 (0,74)	-0,79 (1,05)	t(218) = -11,165 p < 0,001
<b>APTP</b>	2,87 (0,95)	3,74 (0,83)	-0,87 (1,10)	t(261) = -12,810 p < 001
<b>APDB</b>	3,05 (0,96)	3,93 (0,71)	-0,89 (1,08)	t(207) = -11,818 p < 001
<b>APDR</b>	2,80 (0,98)	3,71 (0,70)	-0,91 (1,10)	t(208) = -11,910 p < 001
<b>AAC</b>	2,99 (0,89)	3,47 (0,76)	-0,49 (1,03)	t(218) = -6,998, p < 001
<b>AATP</b>	2,80 (0,91)	3,39 (0,78)	-0,59 (1,00)	t(222) = -8,842 p < 001
<b>AADB</b>	3,09 (0,85)	3,52 (0,73)	-0,43 (0,99)	t(235) = -6,681 p < 001
<b>AADR</b>	3,05 (0,91)	3,46 (0,74)	-0,41 (0,98)	t(261) = -6,710 p < 001

Für alle Gruppen gilt, dass die Confidence in den Forecast des Algorithmus signifikant höher ist als die in den eigenen Forecast. Eine grundsätzliche Stärkung der Confidence durch ein „do your best“-Ziel ist durch den jeweils höchsten Wert über alle Gruppen tendenziell erkennbar.

Die in Abschnitt 4.4 folgenden Hypothesentests zielen darauf ab, die Algorithmen-Aversion und eine potenzielle Beeinflussung durch die eingeführten Variablen zu

<sup>509</sup> Dieser Effekt trat ebenfalls in den Studien von Dietvorst et al. 2015, S. 119 auf: „Moreover, the model's forecasts were much more highly correlated with realized outcomes than were humans' forecasts“.

<sup>510</sup> Dies kann beispielhaft anhand der größten Differenzen sowohl für APDB zu APDR (t(415) = 1,223, p = 0,222), als auch für APC zu APDB (t(425) = -1,859, p = 0,064) gezeigt werden.

ermitteln. Die Verteilung der Entscheidung zwischen dem eigenen Forecast und dem Forecast des Algorithmus in Tabelle 15 stellt die Basis für diesen Hypothesentest dar.

Tabelle 15: Entscheidung im incentivierten Forecast in Studie 1

Gruppe	Wahl auf Forecaster %	Wahl auf Algorithmus in %
APC	33	67
APTP	31	69
APDB	25	75
APDR	32	68
AAC	53	47
AATP	44	56
AADB	48	52
AADR	52	48

Für die AP-Gruppen zeigt sich, dass es in Summe eine starke Präferenz für den Algorithmus gibt. Mit 75 % fällt dieser Wert in der APDB-Gruppe am größten aus. Für die AA-Gruppen zeigt sich ein weniger homogenes Bild. In Summe entscheiden sich die Teilnehmer deutlich häufiger für den eigenen Forecast. Es fällt dabei der Wert in der AATP-Gruppe mit 56 % etwas ab.

## 4.4 Hypothesentests und Beurteilung der empirischen Ergebnisse

### 4.4.1 Hypothesentests

Für die folgenden Hypothesentests wird ein einseitiger Chi-Quadrat-Test durchgeführt.<sup>511</sup> Es wird dabei jeweils die Verteilung zwischen der Wahl für den Forecast des Algorithmus bzw. für den eigenen Forecast herangezogen. Die jeweils untersuchten Gruppen referenzieren auf die Nomenklatur in Tabelle 5.

Für H1 wird die Verteilung zwischen APC und AAC untersucht. Die Algorithmen-Aversion ist hier signifikant und H1 wird angenommen (siehe Tabelle 16). Weitere Interpretationen und Auswertungen zu H1 finden sich in Abschnitt 4.4.2. Tabelle 16 zeigt, wie sich die Algorithmen-Aversion innerhalb der Gruppen mit Leistungsfeedback – also der Algorithmen-Aversion-Gruppen – verhält. Es wird dazu getestet, inwiefern eine Beeinflussung durch die Treatments gegenüber der AAC erfolgt ist. Dabei

<sup>511</sup> Siehe Abschnitt 3.3.

## Studie 1

wird jeweils die AAC mit der AATP (H2), der AADB (H3) und der AADR (H4) auf die Reduktion in der Algorithmen-Aversion getestet. Tabelle 16 zeigt, dass die Einflussnahme von Leistungsfeedback ein robuster Effekt ist. Jedoch ist die Algorithmen-Aversion unter Zeitdruck signifikant niedriger als in der Kontrollgruppe. H2 wird damit angenommen. Die Hypothesen H3 und H4 müssen abgelehnt werden. Tabelle 17 bis Tabelle 20 zeigen die Kreuztabellen, auf denen die Hypothesentests beruhen.

Tabelle 16: Hypothesentests Studie 1

Hypothese	$\chi^2$	df	p	n
H1	18,847	1	< 0,001 <sup>512</sup>	438
H2	3,606	1	0,029 <sup>513</sup>	442
H3	1,503	1	0,110	456
H4	0,062	1	0,402	481

Tabelle 17: Kreuztabelle H1

		Experimentalgruppe		
		APC	AAC	Zeilensumme
Entscheidung	Wahl auf Forecaster	72 (32,9 %)	117 (53,4 %)	189 (43,2 %)
	Wahl auf Algorithmus	147 (67,1 %)	102 (46,6 %)	249 (56,8 %)
Spaltensumme		219 (100 %)	219 (100 %)	438 (100 %)

Tabelle 18: Kreuztabelle H2

		Experimentalgruppe		
		AAC	AATP	Zeilensumme
Entscheidung	Wahl auf Forecaster	117 (53,4 %)	99 (44,4 %)	216 (51,1 %)
	Wahl auf Algorithmus	102 (46,6 %)	124 (55,6 %)	226 (48,9 %)
Spaltensumme		219 (100 %)	223 (100 %)	442 (100 %)

<sup>512</sup> Effektstärke von H1 mit n = 438, wird mit Cramer's-V = 0,207 angegeben. Es liegt demnach ein kleiner bis mittlerer Effekt vor.

<sup>513</sup> Effektstärke von H2 mit n = 442, wird mit Cramer's-V = 0,090 angegeben. Es liegt demnach ein kleiner Effekt vor.

## Studie 1

Tabelle 19: Kreuztabelle H3

		Experimentalgruppe		
		AAC	AADB	Zeilensumme
Entscheidung	Wahl auf Forecaster	117 (53,4 %)	113 (47,7 %)	230 (50,4 %)
	Wahl auf Algorithmus	102 (46,6 %)	124 (52,3 %)	226 (49,6 %)
Spaltensumme		219 (100 %)	237 (100 %)	456 (100 %)

Tabelle 20: Kreuztabelle H4

		Experimentalgruppe		
		AAC	AADR	Zeilensumme
Entscheidung	Wahl auf Forecaster	117 (53,4 %)	137 (52,3 %)	254 (52,8 %)
	Wahl auf Algorithmus	102 (46,6 %)	125 (47,7 %)	227 (47,2 %)
Spaltensumme		219 (100 %)	262 (100 %)	481 (100 %)

### 4.4.2 Zusatzanalysen und Diskussion H1

Die Zusatzanalysen und die Diskussion in diesem Abschnitt ordnen die Ergebnisse von H1 in den Kontext von Dietvorst et al. (2015) ein. Die Algorithmen-Aversion ist in der vorliegenden Untersuchung signifikant und H1 wurde angenommen.

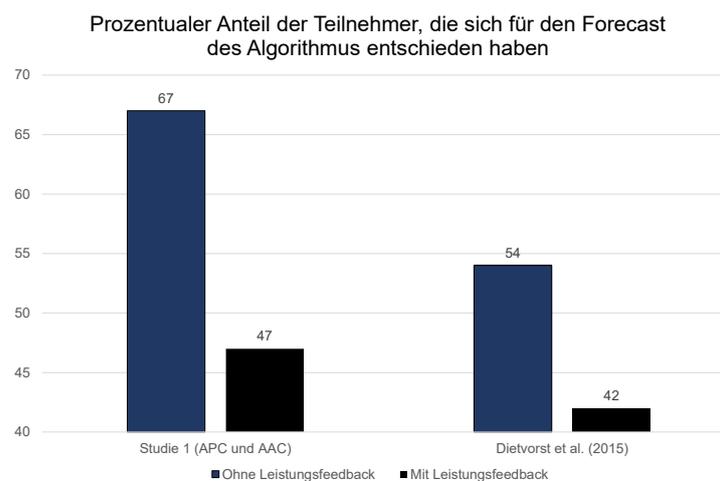


Abbildung 38: Algorithmen-Aversion im Vergleich zu Dietvorst et al. (2015)<sup>514</sup>

<sup>514</sup> Eigene Darstellung.

Verglichen mit der Studie von Dietvorst et al. (2015) ist das Delta zwischen Algorithmen-Appreciation- und Algorithmen-Aversions-Gruppe von 12 % auf 19 % gestiegen (siehe Abbildung 38). Dabei sei gesagt, dass die Betrachtung der Algorithmen-Aversion-Gruppen zwischen dieser Arbeit und Dietvorst et al. (2015) zeigt, dass es einen Anstieg der Algorithmen-Aversion von 42 % auf 47 % gab.<sup>515</sup> Dieser Anstieg, gemessen durch den Vergleich dieser Arbeit mit Dietvorst et al. (2015), ist jedoch mit 5 % nicht signifikant ( $\chi^2(1) = 1,199$ ,  $p = 0,273$ ,  $n = 474$ ). Es kann demnach interpretiert werden, dass die gestiegene Algorithmen-Aversion im Wesentlichen auf die gestiegene Algorithmen-Appreciation zurückzuführen ist.

Der in Abschnitt 2.1.3 beschriebene Mediationseffekt auf die Algorithmen-Aversion durch den Verlust der Confidence in den Forecast des Algorithmus wird in der vorliegenden Untersuchung bestätigt. Eine genauere Betrachtung der Daten aus Tabelle 14 mit der Confidence in den Forecast des Algorithmus und in den eigenen Forecast zeigt, dass die Confidence in den Algorithmus durch das Leistungsfeedback signifikant abnimmt (APC zu AAC  $t(436) = 4,330$ ,  $p < 0,001$ ), während die Confidence in den eigenen Forecast durch das Feedback nicht beeinflusst wird ( $t(436) = 0,051$ ,  $p = 0,959$ ).<sup>516</sup> Dementsprechend wird der Mediationseffekt der Confidence in den Algorithmus durch eine Mediationsanalyse bestätigt. Hierzu wird eine binäre Mediationsanalyse mit 95 % Konfidenzintervallen um den indirekten Effekt gerechnet (siehe Abbildung 39). Die abhängige Variable ist die Entscheidung im incentivierten Forecast (Y),<sup>517</sup> die unabhängige Variable ist, ob Leistungsfeedback erhalten wurde (X).<sup>518</sup> Der Mediator ist die Confidence in den Forecast des Algorithmus (M).<sup>519</sup>

Mit dem indirekten Effekt  $ab = -0,2214$ , 95% CI  $[-0,3859, -0,0971]$ ,  $n = 438$ , mediiert die Confidence in den Algorithmus die Algorithmen-Aversion. Für den a-Pfad gilt ein Koeffizient von  $-0,3105$  mit  $p < 0,001$ , für den b-Pfad gilt  $0,7132$  mit  $p < 0,001$ . Für

---

<sup>515</sup> Dietvorst et al. 2015, S. 120 bezeichnen diese Gruppe als „Control“ und „Saw Results of Model’s and Human’s Forecasts“.

<sup>516</sup> Dies bestätigt die Ergebnisse von Dietvorst et al. (2015) mit der Veränderung der Confidence in den Algorithmus durch Leistungsfeedback, mit  $t(519) = 6,182$ ,  $p < 0,001$ , und der Confidence den eigenen Forecast, mit  $t(518) = 0,322$ ,  $p = 0,748$ .

<sup>517</sup> Kodiert mit: 0 = Wahl auf eigenen Forecast, 1 = Wahl auf Algorithmus.

<sup>518</sup> Kodiert mit: 0 = ohne Leistungsfeedback, 1 = mit Leistungsfeedback.

<sup>519</sup> Exemplarisch kann für den nicht signifikanten Zusammenhang von Leistungsfeedback und der Confidence in den eigenen Forecast gezeigt werden, dass hier keine Mediation vorliegt: Mit dem indirekten Effekt  $ab = 0,0042$ , 95 % CI  $[-0,1579, 0,1666]$ ,  $n = 438$ , hat die Confidence in den eigenen Forecast keine Auswirkungen auf die Algorithmen-Aversion. Für den a-Pfad gilt ein Koeffizient von  $-0,0046$  mit  $p = 0,959$ , für den b-Pfad gilt  $0,9158$  mit  $p < 0,001$ . Für den direkten Effekt gilt, wenn auf die Confidence in den eigenen Forecast kontrolliert wird,  $c' = 0,9910$ ,  $p < 0,001$ .

## Studie 1

den direkten Effekt gilt, wenn auf die Confidence in den Algorithmus kontrolliert wird,  $c' = -0,6841$ ,  $p < 0,001$ .

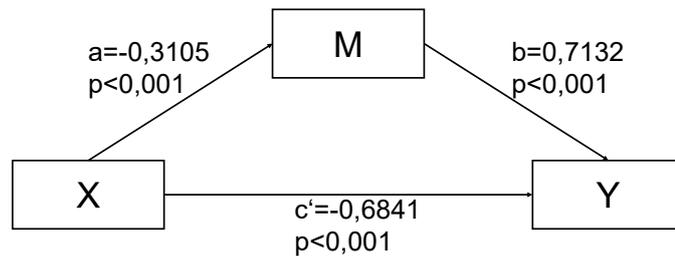


Abbildung 39: Mediationsanalyse zur Confidence in den Algorithmus in H1<sup>520</sup>

Da die größere Algorithmen-Aversion auf einer größeren Algorithmen-Appreciation beruht, wird die APC hier nochmals detailliert untersucht. Die gestiegene Algorithmen-Appreciation kann durch den Vergleich der APC und der Algorithmen-Appreciation-Gruppe 3b aus Dietvorst et al. (2015) erfolgen.<sup>521</sup> Hier zeigt ein Chi-Quadrat-Test, dass es einen signifikanten Anstieg von 2014 bis 2020 gibt:  $\chi^2(1) = 8,240$ ,  $p = 0,004$ ,  $n = 486$ . Der Prozentsatz der Teilnehmer, die sich ohne Kenntnisstand zur Leistung und Präzision eines Algorithmus auf dessen Performance verlassen, stieg signifikant von 54 % in Dietvorst et al. (2015) auf 67 % in der vorliegenden Untersuchung an. Demnach ist eine Tendenz erkennbar, dass die Algorithmen-Appreciation, also das Vertrauen in unbekannte Algorithmen, im Zeitverlauf zunimmt.<sup>522</sup>

Da Algorithmen im privaten wie auch beruflichen Alltag immer mehr zum Einsatz kommen, gewöhnen sich Forecaster zunehmend an die Arbeit mit ihnen.<sup>523</sup> Es ist für viele Forecaster daher zur Routine geworden, sich auf die Algorithmen zu verlassen, was dazu führt, dass sie Confidence in diese aufbauen, selbst wenn sie keinerlei Leistungshinweise zu einem spezifischen Algorithmus haben. Da jedoch einerseits auch gute Algorithmen versagen können und andererseits unbekannte Algorithmen eine sehr schlechte Leistung aufweisen können, sollte die Funktion und Leistung hinterfragt

<sup>520</sup> Eigene Darstellung.

<sup>521</sup> Dietvorst et al. 2015, S. 120 bezeichnen diese Gruppe als „Control“ und „Saw Results of Model's and Human's Forecasts“.

<sup>522</sup> Der Vergleich von Experimentalgruppen aus unterschiedlichen Erhebungen unterliegt einer eingeschränkten Aussagekraft. Siehe dazu Abschnitt 6.2.1.

<sup>523</sup> Vgl. Griffin und Wright 2015, S. 377; Quattrone 2016, S. 118.

werden, um negative Konsequenzen schlecht funktionierender Algorithmen zu vermeiden.<sup>524</sup> Auch wenn die Forecaster in dieser Untersuchung einen größeren Bonus erhielten, wenn sie sich auf den Algorithmus verließen, sollte aufgrund dieses potenziellen Versagens mehr Skepsis gegenüber unbekanntem Algorithmen herrschen.<sup>525</sup>

Die Ergebnisse zur Algorithmen-Aversion zeigen, dass Forecaster bei Vorhandensein von Leistungsfeedback eine große Intoleranz gegenüber unvollkommenen Algorithmen haben. Sobald sie wahrnehmen, dass der Algorithmus Fehler macht, lehnen sie den Algorithmus zugunsten ihrer schlechteren Vorhersageleistung ab. Es wird in Abschnitt 4.3.2.2 gezeigt, dass die Ablehnung von Algorithmen zu einer geringeren Vorhersagegenauigkeit und damit zu einer kostspieligen Algorithmen-Aversion führt. Diese Kostspieligkeit bezieht sich auf den entgangenen Bonus der Teilnehmer. Mit Fokus auf die negativen Effekte des Leistungsfeedbacks sollten Unternehmen sicherstellen, dass ihre Algorithmen eine akzeptable Leistung erbringen, wenn sie eingeführt werden. Die Einführung von schlechten Algorithmen führt zu einer Algorithmen-Aversion, die nur schwer wieder abgebaut werden kann. Es sollte daher ein besonderes Augenmerk auf den Zeitpunkt der Einführung gelegt werden. Algorithmen dürfen erst dann für Forecasts zur Verfügung gestellt werden, wenn eine sehr hohe Präzision erreicht werden kann. Weiter kann kommuniziert werden, dass ein Algorithmus lernfähig ist<sup>526</sup>, oder er für Regelprozesse mit sehr häufiger Interaktion eingeführt wird.<sup>527</sup> Für die einmalige Verwendung von Algorithmen hingegen werden aus Sicht der Akzeptanz keine Einschränkungen gegeben.

### 4.4.3 Algorithmen-Aversion H2 bis H4

Abbildung 40 unterstützt die Hypothesentests in Tabelle 16. Es ist dabei ersichtlich, wie die Wahl zugunsten des Algorithmus mit 56 % unter Zeitdruck gegenüber 47 % in der Kontrollgruppe hoch ausfällt.

---

<sup>524</sup> Vgl. Fildes et al. 2009, S. 3.

<sup>525</sup> Vgl. McKinsey 2017, S. 24.

<sup>526</sup> Vgl. Berger et al. 2021, S. 55.

<sup>527</sup> Vgl. Filiz et al. 2021, o. S.

## Studie 1

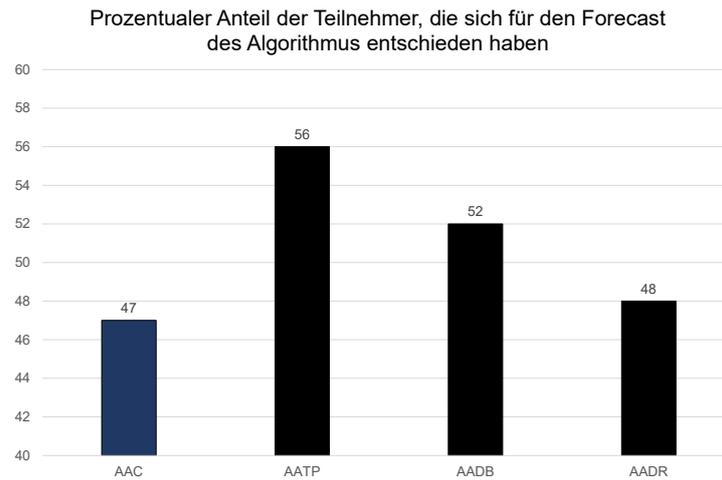


Abbildung 40: Reduktion der Algorithmen-Aversion H2–H4 gegenüber H1<sup>528</sup>

Um den Einfluss der drei Variablen Zeitdruck, „do your best“-Ziel und Entscheidungsrechte besser zu erklären, zeigt Abbildung 41, wie sich die Algorithmen-Aversion innerhalb der Treatments verhält. Hier werden die APTP mit der AATP, die APDB mit der AADB und die APDR mit der AADR gegenübergestellt.

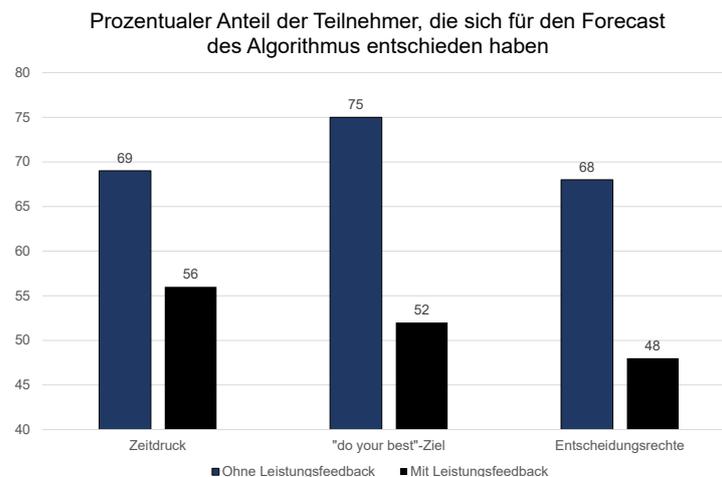


Abbildung 41: Algorithmen-Aversion innerhalb der Treatmentgruppen<sup>529</sup>

Analog zu Abbildung 41 unterstreicht Tabelle 21 die Robustheit der Algorithmen-Aversion aus dem vorhergehenden Abschnitt. Durch Leistungsfeedback entsteht eine signifikante Algorithmen-Aversion über alle drei Treatmentgruppen hinweg. Die hohe Algorithmen-Appreciation in der APDB-Gruppe bewirkt hier der Effekt, dass die Aversion besonders stark ausfällt, wenn die Algorithmen-Appreciation stark ist. Die geringste Algorithmen-Aversion herrscht unter Zeitdruck.

<sup>528</sup> Eigene Darstellung.

<sup>529</sup> Eigene Darstellung.

## Studie 1

Tabelle 21: Algorithmen-Aversion innerhalb der Treatmentgruppen in Studie 1

Gruppen	$\chi^2$ <sup>530</sup>	df	p	n
<b>APTP zu AATP</b>	9,566	1	0,002	486
<b>APDB zu AADB</b>	24,422	1	< 0,001	445
<b>APDR zu AADR</b>	20,335	1	< 0,001	471

Von besonderem Interesse ist die Abschwächung der Algorithmen-Aversion durch Zeitdruck und die Frage, weshalb die anderen beiden Treatments keine Wirkung gezeigt haben. Um dies zu erklären, werden analog zu Abschnitt 2.1.3 die Confidencewerte herangezogen. Ein Vergleich der Confidencewerte mittels t-Tests zeigt ein differenziertes Bild zwischen den Treatmentgruppen im Vergleich zur AAC (siehe Tabelle 22).

Tabelle 22: Confidencewerte innerhalb der Algorithmen-Aversion-Gruppen in Studie 1 (M)

Confidence Algorithmus	Differenz zur AAC	Ungepaarter t-Test
<b>AATP</b>	-0,08	t(440) = -1,094, p = 0,275
<b>AADB</b>	0,04	t(454) = 0,629, p = 0,529
<b>AA DR</b>	-0,01	t(479) = -0,191, p = 0,849
<b>Confidence Forecaster</b>		
<b>AATP</b>	-0,18	t(440) = -2,151, p = 0,032
<b>AADB</b>	0,10	t(454) = 1,257, p = 0,209
<b>AA DR</b>	0,07	t(479) = 0,815, p = 0,416

Es lassen sich über alle drei Treatmentgruppen hinweg keine signifikanten Veränderungen der Confidence in den Algorithmus gegenüber der AAC feststellen. Es kann interpretiert werden, dass die Treatments keinen Einfluss darauf haben, wie der Algorithmus wahrgenommen wird. Dies war für Zeitdruck so zu erwarten, denn der Zeitdruck wirkt nur auf die Forecaster und nicht auf den Algorithmus. Für das „do your best“-Ziel gilt dies ebenfalls. Für die Entscheidungsrechte und den damit verbundenen Eingriff in die Daten, mit denen der Algorithmus arbeitet, zeigt sich, dass hieraus keine verstärkte Confidence in den Algorithmus resultiert. Dies wird in Abschnitt 4.4.3.2 weiter untersucht und erklärt.

<sup>530</sup> Da hierzu keine Hypothesen aufgestellt wurden, wird ein zweiseitiger Chi-Quadrat-Test gerechnet.

Für das „do your best“-Ziel und die Entscheidungsrechte ergibt sich keine Änderung der Confidence in den eigenen Forecast. Jedoch zeigt sich, dass die Confidence in den eigenen Forecast durch Zeitdruck signifikant verringert wird. Basierend auf den Confidencewerten in Tabelle 22 werden im Folgenden die drei Hypothesen nochmals näher diskutiert.

### 4.4.3.1 Zusatzanalysen und Diskussion H2

H2 fokussiert die Zeit, die den Forecastern für einen Forecast zur Verfügung steht. Durch diesen Zeitdruck wirkt auf Forecaster zusätzlicher Stress in der Erstellung ihrer Forecasts. Es wurde unter Zeitdruck eine geringere Algorithmen-Aversion erwartet. Aus der Literatur wurde der Schluss gezogen, dass die Einführung von Zeitdruck die Confidence der Forecaster in ihre eigenen Forecasts verringert. Der Zeitdruck sollte keinen Einfluss auf die Art und Weise haben, wie die Forecaster die Leistung des Algorithmus bewerten.

Tabelle 22 zeigt, dass die Confidence in den eigenen Forecast der Forecaster unter Zeitdruck signifikant niedriger ist als in der Kontrollbedingung. Der Zeitdruck führt demnach dazu, dass sich die Forecaster in Bezug auf ihre eigenen Forecasts zunehmend unsicherer fühlen. Die Wirkung der Veränderung der Confidence in den eigenen Forecast auf die Algorithmen-Aversion wurde in der Literatur bislang nicht gezeigt bzw. wahrgenommen. Da Zeitdruck die Algorithmen-Aversion und gleichzeitig die Confidence in den eigenen Forecast absenkt, wird dieser Zusammenhang näher untersucht.

Es wird hierzu getestet, inwiefern die Zusammenhänge der Mediation durch die Confidencewerte gelten.<sup>531</sup> Hierzu wird eine binäre Mediationsanalyse mit 95 % Konfidenzintervallen um den indirekten Effekt gerechnet. Die Stichprobengröße beträgt 442 Teilnehmer. Dies umfasst die AAC und die AATP. Die abhängige Variable ist die Entscheidung des Teilnehmers im incentivierten Forecast für oder gegen den Algorithmus (Y).<sup>532</sup> Die unabhängige Variable ist, ob er unter Zeitdruck stand oder nicht (X).<sup>533</sup> Der Mediator ist die Confidence in den eigenen Forecast (M).

---

<sup>531</sup> Siehe dazu Abschnitt 2.1.3.

<sup>532</sup> Kodiert mit: 0 = Wahl auf eigenen Forecast, 1 = Wahl auf Algorithmus.

<sup>533</sup> Kodiert mit: 0 = ohne Zeitdruck, 1 = mit Zeitdruck.

## Studie 1

Mit dem indirekten Effekt  $ab = 0,1515$ , 95 % CI  $[0,0142, 0,3202]$ ,  $n = 442$ , mediert die Confidence in den eigenen Forecast die Algorithmen-Aversion, wenn Forecaster unter Zeitdruck stehen. Für den a-Pfad gilt ein Koeffizient von  $-0,1836$  mit  $p = 0,032$ , für den b-Pfad gilt  $-0,8251$  mit  $p < 0,001$ . Für den direkten Effekt gilt, wenn auf die Confidence in den eigenen Forecast kontrolliert wird,  $c' = 0,2539$ ,  $p = 0,211$ .

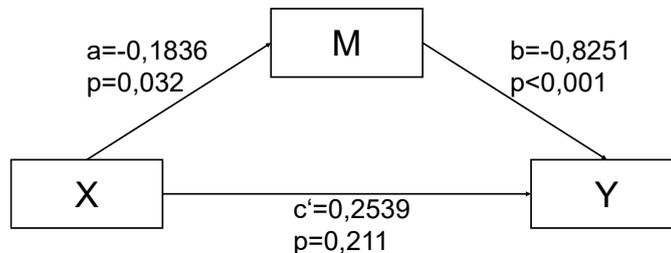


Abbildung 42: Mediationsanalyse Zeitdruck<sup>534</sup>

Abbildung 42 zeigt den Einfluss von Zeitdruck auf die Algorithmen-Aversion. Diese Erkenntnisse zum Zeitdruck geben eine wesentliche Weiterentwicklung im Gesamtbild der Algorithmen-Aversion wieder. Bislang besagt die Literatur, dass die Aversion gegenüber Algorithmen aus der Reduktion der Confidence in den Algorithmus durch Leistungsfeedback resultiert. Dabei spielt die Confidence in den eigenen Forecast keine nennenswerte Rolle. Dies muss durch die Erkenntnisse dieser Untersuchung erweitert werden: Ist die Confidence in den Algorithmus bereits durch Leistungsfeedback reduziert, kann die resultierende Aversion wiederum durch die Reduzierung der Confidence in den eigenen Forecast reduziert werden.

Darauf aufbauend ergibt sich für die Literatur zum Zeitdruck eine Ergänzung. Die empirischen Daten zeigen, dass sich die Forecaster entsprechend der prognostizierten Wirkung von Zeitdruck verhalten. Zeitdruck wird häufig als belastende Arbeitsanforderung wahrgenommen<sup>535</sup>, die dazu führt, dass sich die Leistung der Forecaster verschlechtert, wenn sie aufgrund von Druck bei einer Aufgabe das Vertrauen in ihre eigenen Fähigkeiten verlieren.<sup>536</sup> Der Verlust der Confidence in das eigene Urteilsvermögen muss jedoch in der Zusammenarbeit mit Algorithmen differenziert betrachtet werden. Forecaster können durch den Algorithmus einen Ausweg aus der Drucksituation finden.

<sup>534</sup> Eigene Darstellung.

<sup>535</sup> Vgl. Mischer et al. 2021, S. 19.

<sup>536</sup> Vgl. Beilock und Carr 2001, S. 701.

### 4.4.3.2 Zusatzanalysen und Diskussion H3

H3 befasst sich mit der Zielsetzung von Forecastern. Anstelle eines spezifischen und schwierigen Ziels wurde den Forecastern ein „do your best“-Ziel gegeben. Durch dieses „do your best“-Ziel wurde erwartet, dass sie weniger Druck in der Aufgabe verspüren und daher eine reflektierte Bewertung der Leistung vornehmen können. Dies könnte zu einer höheren Confidence in den Algorithmus führen.

Die Teilnehmer in der DB-Bedingung haben in Tabelle 14 die höchsten Confidencewerte hinsichtlich der Confidence in den eigenen Forecast und in den des Algorithmus. Es ist also davon auszugehen, dass das Treatment von den Teilnehmern angenommen wurde und wirksam war. Gleichzeitig haben die Teilnehmer den höchsten Wert in der Algorithmen-Appreciation- und den zweithöchsten in der Algorithmen-Aversion-Gruppe. Diese Werte sind jedoch nicht statistisch signifikant und so zeigt sich, dass sich die Zuweisung eines „do your best“-Ziels nicht nachhaltig auf die Algorithmen-Aversion der Forecaster auswirkt und entsprechend H3 abgelehnt werden muss.

In Übereinstimmung mancher früherer Ergebnissen zu „do your best“-Zielen<sup>537</sup> verhielten sich die Teilnehmer in dieser Bedingung wie die Teilnehmer in der Kontrollbedingung. Auch, wenn sie ein breiteres Spektrum an akzeptablen Ergebnissen für den Forecast wahrnehmen, ändern sie ihr Verhalten gegenüber dem Einsatz von Algorithmen nicht. Es scheint, als entstehe durch das „do your best“-Ziel kein Drang, den Algorithmus ausgiebiger zu testen oder ihn aus Neugier zu nutzen. Es gibt darüber hinaus einen bedeutsamen Unterschied gegenüber der Kontrollgruppe in Bezug auf die praktischen Auswirkungen: Aufgrund der fixen Incentivierung verdienten sie bei ähnlicher Algorithmen-Aversion wie die Kontrollbedingung die mit Abstand höchsten Bonuszahlungen, weshalb von einem „do your best“-Ziel im Forecasting mit Algorithmen Abstand genommen werden sollte.

### 4.4.3.3 Zusatzanalysen und Diskussion H4

Die Forecaster in H4 wählten die Inputfaktoren für den Algorithmus vor den zehn Übungsforecasts. Anschließend gab es die Möglichkeit, die Auswahl der Inputfaktoren nach den zehn Übungsforecasts, also vor dem incentivierten Forecast, nochmals zu

---

<sup>537</sup> Vgl. Akın und Karagözoğlu 2017, S. 193.

## Studie 1

wechseln. Aus den 262 Teilnehmern dieser Gruppe (AADR) haben sich 30 dazu entschieden, die Inputfaktoren nach den Übungsforecasts zu wechseln, 232 sind bei ihrer ursprünglichen Auswahl geblieben. Diese geringe Anzahl von rund 13 %, die sich für eine Änderung der Inputfaktoren entschieden hat, lässt vermuten, dass durch den Wechsel von einem der vier Inputfaktoren keine Verbesserung der Performance des Algorithmus erwartet wurde. Tabelle 23 zeigt, dass es tendenziell die Teilnehmer mit einer höheren Confidence in den eigenen Forecast sind, die auch den Algorithmus nochmals verändern:

*Tabelle 23: Änderungen nach Übungsforecasts an der Auswahl der Eingabeinformationen (M und SD)*

<b>Gruppe</b>	<b>Änderungen vorgenommen (n = 30)</b>	<b>Keine Änderungen vorgenommen (n = 232)</b>	<b>Differenz</b>	<b>Ungepaarter t-Test</b>
<b>Confidence Algorithmus</b>	3,67 (0,80)	3,44 (0,72)	0,23	t(260) = 1,627 p = 0,105
<b>Confidence Forecaster</b>	3,37 (1,13)	3,01 (0,88)	0,35	t(260) <sup>538</sup> = 1,653 p = 0,108

Da weder signifikante Unterschiede in der Confidence in den eigenen Forecast noch in der Confidence in den Algorithmus vorliegen, liegt auch keine Änderung der Algorithmen-Aversion vor. Aus der Gruppe der 30 Teilnehmer, die eine Veränderung an den Inputfaktoren des Algorithmus vorgenommen haben, haben sich 15 für und 15 gegen den Algorithmus entschieden. Aus den anderen 232 Teilnehmern wählen 110 Teilnehmer den Algorithmus und 122 Teilnehmer den eigenen Forecast. Ein zweiseitiger Chi-Quadrat-Test ergibt demnach hier keine Unterschiede der Algorithmen-Aversion innerhalb der AADR Gruppe: ( $\chi^2(1) = 0,071$ ,  $p = 0,790$ ,  $n = 262$ ).

Zwei mögliche Erklärungen hierfür sind denkbar: Die Teilnehmer sahen sich nicht in der Lage, mittels der Auswahl von vier aus fünf Inputfaktoren den Algorithmus zu verbessern. Gelöst werden könnte dies in künftigen Untersuchungen durch die Möglichkeit eines stärkeren Eingriffs in die Inputfaktoren oder durch eine bessere Aufklärung über die Funktionsweise des Algorithmus. Dadurch könnte die potenzielle Senkung der Algorithmen-Aversion durch die Entscheidungsrechte erreicht werden. Weiter ist es möglich, dass die Teilnehmer nicht überzeugt von ihren eigenen Fähigkeiten waren. Denn um eine Notwendigkeit der Veränderung der Inputfaktoren zu erkennen,

<sup>538</sup> Varianzen sind nicht gleich, weshalb ein Welch-Test verwendet wurde. Siehe dazu auch: Rasch et al. 2011.

bedarf es hinreichender Sicherheit darüber, welche Inputfaktoren ausschlaggebend für bessere Forecasts sind. Es ist also möglich, dass Teilnehmer, die sich besonders gut mit den Daten auseinandergesetzt haben, zwar in der Lage waren, eine sinnvolle Veränderung an den Inputfaktoren vorzunehmen, jedoch ebenso einen starken Lerneffekt hatten. Dadurch wurde die Änderung der Eingabeinformationen nicht mehr für nötig empfunden, da sie ohnehin ihren eigenen Forecast wählen wollten. Für diese These spricht eine gute Forecastinggenauigkeit in Tabelle 12.

#### 4.4.4 Schlussfolgerungen

Tabelle 24 fasst die Hypothesen, die prognostizierte Wirkung und den jeweilige Hypothesentest zusammen. Durch H1 wird die Algorithmen-Aversion untersucht und nachgewiesen. Da Forecaster zunehmend an die Arbeit mit Algorithmen gewöhnt sind, ist es für viele Forecaster zur Routine geworden, Algorithmen in Arbeitsabläufe zu integrieren. Dies führt zu einer großen Algorithmen-Appreciation: Algorithmen werden bereitwillig angenommen, wenn deren Leistung und Präzision nicht bekannt sind.<sup>539</sup> Ein Vergleich mit den Daten aus Dietvorst et al. (2015) zeigt, dass insbesondere die Algorithmen-Appreciation über die Zeit anzusteigen scheint. Dies ist insofern bedenklich, da die Forecaster kein Wissen über die Struktur und Zuverlässigkeit des Algorithmus haben und ihm somit eine gewisse Skepsis entgegengebracht werden sollte.<sup>540</sup> Die Algorithmen-Aversion führt zu einer geringeren Genauigkeit der Forecasts und resultiert damit in Verlusten, der leistungsabhängigen Vergütung beim Forecaster.

Tabelle 24: Zusammenfassung der Hypothesentests in Studie 1

Hypothese	Prognostizierte Wirkung	Hypothesentest
H1: Durch Leistungsfeedback entsteht eine Algorithmen-Aversion.	↑	<input checked="" type="checkbox"/>
H2: Zeitdruck reduziert die Algorithmen-Aversion.	↓	<input checked="" type="checkbox"/>
H3: Ein „do your best“-Ziel reduziert die Algorithmen-Aversion.	↓	<input type="checkbox"/>
H4: Entscheidungsrechte für Eingabeinformationen, die der Algorithmus verarbeitet, reduzieren die Algorithmen-Aversion.	↓	<input type="checkbox"/>

<sup>539</sup> Vgl. Griffin und Wright 2015, S. 377; Quattrone 2016, S. 118.

<sup>540</sup> Vgl. Fildes et al. 2009, S. 3; McKinney et al. 2017, S. 63.

## Studie 1

Mit H2 konnte gezeigt werden, dass die Algorithmen-Aversion unter Zeitdruck abgeschwächt wird. Dies ist darauf zurückzuführen, dass unter Zeitdruck die Confidence in den eigenen Forecast sinkt. Wenn die Confidence in den eigenen Forecast abnimmt, sinkt auch die Algorithmen-Aversion. Studie 1 ergänzt damit die Literatur zum Thema *Zeitdruck und Algorithmen-Aversion*. In der Regel wird Zeitdruck als etwas Negatives wahrgenommen, weil er zu Unsicherheit und teilweise gar zu Angst führt.<sup>541</sup> Dies wiederum führt zu schlechter Leistung oder eingeschränkter Kreativität beim Finden neuer Lösungen und Arbeitsweisen. Durch die Kombination zweier Stränge der Literatur – zur Algorithmen-Aversion und zum Zeitdruck – ergibt sich eine Erweiterung zum negativen Effekt von Zeitdruck. Der Verlust der Confidence in den eigenen Forecast kann kompensiert werden durch den Algorithmus. Der Algorithmus hat keine technischen Nachteile und stellt damit eine Arbeitsstrategie dar, die einen Ausweg aus der Drucksituation ermöglicht.

Dies ist einerseits bedeutsam für die Erkenntnis, dass Zeitdruck nicht als per se schlecht anzusehen ist,<sup>542</sup> insbesondere dann nicht, wenn eine Technologie zur Verfügung steht, die unterstützend wirken kann. Diesem Gedanken folgend sind Algorithmen ein Ausweg aus Druck- bzw. Stresssituationen und können insbesondere solchen Forecastern helfen, die ohne den Algorithmus vermehrt Schwierigkeiten in der Aufgabenerfüllung haben. Zeitdruck ist des Weiteren eine Möglichkeit, wie ein Algorithmus in einer bekannten Aufgabe verstärkt eingesetzt werden kann, ohne dass ein Mensch in den Prozess eingreift und damit die Vorhersage des Algorithmus verschlechtert.<sup>543</sup> Dies stellt eine Erweiterung zu den bestehenden Möglichkeiten der Aversionsminderung durch einen Eingriff in das Ergebnis eines Algorithmus dar.

Zeitdruck kann als aktives Steuerungsinstrument zur Reduktion der Algorithmen-Aversion im Forecasting eingesetzt werden. Hierzu können Daten vom Management zurückgehalten oder die Fristen für die Erstellung eines Forecasts verkürzt werden. In beiden Fällen ist eine sinkende Algorithmen-Aversion zu erwarten. Da jedoch dennoch nicht alle Forecaster den Algorithmus wählen und damit unter starken Druck geraten, ist hier von einem zu hohen Zeitdruck abzusehen.<sup>544</sup> Darüber hinaus kann aus Steuerungssicht mehr Toleranz gegenüber Zeitdruck folgen. Das heißt, dem einer

---

<sup>541</sup> Vgl. Pietsch und Messier 2017, S. 51; Wegier und Spaniol 2015.

<sup>542</sup> Vgl. Kirchler et al. 2017, S. 57.

<sup>543</sup> Vgl. Dietvorst et al. 2018, S. 1155; Fildes et al. 2009, S. 3.

<sup>544</sup> Vgl. Kirchler et al. 2017, S. 57.

Forecastingaufgabe inhärenten Zeitdruck muss nicht zwangsläufig durch besseres Zeitmanagement entgegengewirkt werden, sofern es die Möglichkeit gibt, Algorithmen einzusetzen.

Negative Effekte von schwierigen und spezifischen Zielen können eine starke Fokussierung auf die Erreichung dieser Ziele durch die Intensivierung der Arbeit mit bestehenden Methoden erzeugen.<sup>545</sup> Um dem entgegenzuwirken, wurde in H3 ein „do your best“-Ziel für Forecasts eingeführt. Es wurde argumentiert, dass ein „do your best“-Ziel Forecastern ermöglicht, die überlegene, aber nicht perfekte Leistung des Algorithmus rationaler zu verarbeiten und damit besser einzuordnen. Dies führt dem Forecaster vor Augen, dass der Algorithmus besser abschneidet und im Forecasting anstelle des eigenen Forecasts eingesetzt werden sollte.<sup>546</sup> Es wurde erwartet, dass dies die Algorithmen-Aversion abschwächt. Die empirischen Ergebnisse können diese Hypothese nicht bestätigen. Bei einem „do your best“-Ziel ist die Tendenz, unbekannte Algorithmen einzusetzen besonders groß. Damit einhergehend besteht eine große Algorithmen-Aversion nach Leistungsfeedback zum Algorithmus. Dazu muss ergänzt werden, dass die mit dem „do your best“-Ziel einhergehende Incentivierung den Forecastern einen großen Bonus ermöglicht hat. Das heißt, die Forecastingergebnisse werden nicht verbessert, die Algorithmen-Aversion besteht weiterhin und die Incentivierung für diese schlechten Forecasts ist besonders hoch. In Konklusion sollte daher von einem „do your best“-Ziel Abstand genommen werden.

In Bezug auf die Entscheidungsrechte der Forecaster wurde die Hypothese aufgestellt, dass die Teilnehmer durch die Beeinflussung der Eingabeinformationen des Algorithmus mehr Confidence in den Algorithmus gewinnen und die Algorithmen-Aversion damit abgeschwächt wird (H4). Die Literatur zeigt, dass die Algorithmen-Aversion reduziert werden kann, indem Forecaster die Forecasts eines Algorithmus verändern dürfen.<sup>547</sup> Die Anpassung des Ergebnisses des Algorithmus führt jedoch dazu, dass die Genauigkeit der Vorhersage abnimmt, weshalb dieser nachträgliche Eingriff in den Forecast häufig kein probates Mittel zur Senkung der Aversion ist.<sup>548</sup> Durch die Einführung von Entscheidungsrechten für die Eingabeinformationen zeigen die empirischen Ergebnisse keine Unterschiede im Verhalten der Forecaster und demnach auch

---

<sup>545</sup> Vgl. Gold et al. 2014, S. 303; Seckler et al. 2017, S. 22.

<sup>546</sup> Vgl. Webb et al. 2013, S. 1433.

<sup>547</sup> Vgl. Dietvorst et al. 2018, S. 1155.

<sup>548</sup> Siehe dazu beispielsweise Dietvorst et al., S. 1155; Goodwin und Fildes 1999, S. 37.

keine Reduktion der Algorithmen-Aversion. Weiter zeigen die Zusatzanalysen, dass es nur einen geringen Bedarf bei den Teilnehmern gab, die Inputfaktoren anzupassen. Es ist demnach möglich, dass die Teilnehmer nicht in der Lage waren, die Relevanz der Inputfaktoren differenziert zu bewerten. Wenn sie selbst nicht in der Lage waren, die Relevanz der Inputfaktoren zu bewerten, können sie auch keine sinnvolle Auswahl für den Algorithmus treffen. In Bezug auf diese Entscheidungsrechte bei der Arbeit mit Algorithmen muss daher festgehalten werden, dass die Algorithmen-Aversion nur dann verringert werden kann, wenn die Forecaster die Ausgabe eines Algorithmus beeinflussen können – nicht aber den Input.

### 4.4.5 Limitationen

Die empirische Untersuchung in Studie 1 unterlag einigen Limitationen, die es bei der Interpretation der Ergebnisse zu berücksichtigen gilt. Aus den Zusatzanalysen zu H1 wurde eine Aussage zum zeitlichen Verlauf der Algorithmen-Aversion getroffen, die es kritisch zu hinterfragen gilt. Bei der Interpretation der Algorithmen-Aversion im Zeitverlauf in Abschnitt 4.4.2 liegt aufgrund der Betrachtung lediglich zweier Zeitpunkte eine stark begrenzte Aussagekraft vor. Es muss hier außerdem erwähnt werden, dass ein Vergleich zwischen den Daten unterschiedlicher Forscher und Erhebungen gemacht wird. Auch wenn die verglichenen Daten aus dem gleichen experimentellen Aufbau hervorgehen, ist es nicht gesichert, dass die gezeigten Unterschiede auf die zeitliche Diskrepanz zurückzuführen sind. Eine regelmäßige Untersuchung der Algorithmen-Aversion im Zeitverlauf kann sicherstellen, dass auch Zeiteffekte berücksichtigt werden.

Eine Algorithmen-Aversion tritt vermindert auf, wenn ein Forecast unter Zeitdruck erstellt wird. Zeitdruck kann als unangenehm empfunden werden und der Algorithmus als Ausweg dienen. In dieser Studie wurde der Zeitdruck auf ein moderates Niveau von zwölf Sekunden für einen Forecast konstant festgelegt. Das reduziert die Aussagekraft zum Zeitdruck eben auf dieses moderate Niveau. Künftige Forschung kann sich einer Varianz im Zeitdruck widmen. Es ist anzunehmen, dass durch größeren Zeitdruck die Algorithmen-Aversion weiter abnimmt. Dabei kann auch die zugrundeliegende Aufgabe verändert werden. Die Algorithmen-Aversion in Aufgaben, die per

se länger dauern oder komplexer sind,<sup>549</sup> könnte noch mehr durch Zeitdruck beeinflusst werden. Hier wird der Nutzen als *Ausweg* aus dem Stress durch Zeitdruck größer, da die Stresssituation mit Zeitdruck länger andauert.

Auf der Grundlage des Modus für die leistungsabhängige Vergütung<sup>550</sup> erhielten die Forecaster in dieser Studie für ihren eigenen Forecast im Durchschnitt einen Bonus von 0,27 \$, während diejenigen, die sich auf den Algorithmus verließen, 0,49 \$ erhielten. Dies kann als Erwartungswert für die Auszahlung des Bonus je nach Entscheidung im incentivierten Forecast verstanden werden. Der Erwartungswert der Auszahlung für die „do your best“-Bedingungen unterschied sich hingegen. Das „do your best“-Ziel impliziert einen festen, nicht an die Leistung der Forecaster gebundenen Bonus. Dadurch erhöhte sich der Erwartungswert auf einen Bonus von 1 \$, unabhängig von der Leistung der Forecaster. Dies war notwendig, um sicherzustellen, dass die Teilnehmer nicht durch kleinere potenzielle Boni demotiviert werden. Im Gegenzug kann nicht ausgeschlossen werden, dass der veränderte Erwartungswert der Auszahlung eine Auswirkung auf das Verhalten der Forecaster hatte.

Um die Forecaster mit Entscheidungsrechten an den Eingabeinformationen auszustatten, konnten sie vier aus fünf Datenpunkten zur Verarbeitung durch den Algorithmus bestimmen. Das heißt, der Algorithmus hat hier weniger Daten zur Verfügung. Die Einschätzung gegenüber dem Algorithmus hat sich dadurch nicht zum Negativen verändert, weshalb dieser Effekt nicht als relevant eingestuft wird (Tabelle 14). Künftige Untersuchungen könnten daran anschließen und untersuchen, inwiefern unterschiedliche Datenmengen, also die Datenbasis des Algorithmus, eine Algorithmen-Aversion beeinflussen können. Berger et al. (2021) vergleichen menschliche Ratschläge mit denen eines Algorithmus und erkennen dabei eine Algorithmen-Aversion.<sup>551</sup> Die Autoren sind der Ansicht, dass der menschliche Entscheidungsprozess ähnlichen Prinzipien folgt, sodass sich menschliche Forecaster hinsichtlich ihrer Entscheidungslogik miteinander identifizieren können. Der Algorithmus hingegen folgt einer Entscheidungslogik und Gewichtungen von Daten im Entscheidungsprozess, die weniger nachvollziehbar erscheint. Menschliche Ratschläge in Entscheidungen sind demnach komplementär zur eigenen Einschätzung. Es ist den Forecastern in dieser

---

<sup>549</sup> Dem Thema *komplexer Aufgaben* in der Algorithmen-Aversion widmet sich eine in 2022 erschienene Studie von Commerford et al. 2022. Siehe darüber hinaus Liu und Li 2012, S. 553; Wood et al. 1990, S. 181.

<sup>550</sup> Siehe Tabelle 2.

<sup>551</sup> Vgl. Berger et al. 2021, S. 58

## Studie 1

Studie zwar möglich, die Variablen, die der Algorithmus verarbeitet, auszuwählen – jedoch nicht die Entscheidungslogik und Gewichtung von Daten aktiv zu beeinflussen. Dies erklärt, weshalb die Bearbeitung eines Outputs eines Algorithmus, wie in Dietvorst et al. (2018), eine komplementäre Interaktion zwischen Forecaster und Algorithmus darstellt.<sup>552</sup> Der Eingriff bei nicht vollständigem Verständnis der Entscheidungslogik des Algorithmus folgt dem jedoch nicht. Damit Entscheidungsrechte die Algorithmen-Aversion absenken können, kann Forecastern vermehrt die Entscheidungslogik des Algorithmus erklärt werden.

Die Stichprobe in dieser Studie besteht nicht aus Forecastingexperten. Entscheidungsrechte für Eingabeinformationen könnten ein effektiverer Weg sein, die Algorithmen-Aversion zu verringern, wenn Experten statt Laien beteiligt sind. Ihre Kompetenz könnte auf den Algorithmus übertragen werden und somit die Confidence in den Algorithmus erhöhen. Dementsprechend können künftige Untersuchungen in Unternehmen unmittelbar bei Forecastingexperten erfolgen. Durch das Framing von Aufgaben bzw. Forecastingaufgaben aus bestimmten Fachbereichen, können solche Untersuchungen auch mit Studenten im jeweiligen Studienfach erfolgen.

Das Versuchsdesign in Studie 1 birgt das Risiko einer Verzerrung der Forecastinggenauigkeit aufgrund der zehn Übungsrunden. Leistungsfeedback kann in diesem Sinne als Training verstanden werden. Die Teilnehmer in den Algorithmen-Aversion-Bedingungen hatten einen Vorteil von zehn Übungsforecasts gegenüber den Teilnehmern, die kein Leistungsfeedback erhalten haben. Es wurde in dieser Arbeit nicht untersucht ob daraus Einflüsse über die Algorithmen-Aversion hinaus entstanden sind. Berger et al. (2021) haben diese Schwäche im Experimentaldesign zur Algorithmen-Aversion ebenfalls identifiziert und alle Teilnehmer die Übungsforecasts ohne Leistungsfeedback zum Algorithmus oder anderen Teilnehmern (Experten) durchführen lassen. Dies hatte keinen Einfluss auf die Algorithmen-Aversion, weshalb zunächst davon auszugehen ist, dass hierdurch keine einschneidende Verzerrung entstanden ist.<sup>553</sup>

Die Betrachtung der drei in diesem Kapitel untersuchten Einflussfaktoren aus dem Arbeitsumfeld von Forecastern, wird im folgenden Kapitel um den Einfluss von Kollegen und damit verbundenem relativem Feedback erweitert. Dabei wird auf das relative

---

<sup>552</sup> Vgl. Dietvorst et al. 2018, S. 1155.

<sup>553</sup> Vgl. Berger et al. 2021, S. 60 und 64.

## Studie 1

Feedback fokussiert, das ein Forecaster erhält, wenn er nicht mehr nur Leistungsfeedback zur eigenen Leistung gegenüber dem Algorithmus, sondern auch gegenüber der Leistung seiner Kollegen erhält.

## 5 Studie 2

Aus dem Stand der Forschung in Kapitel 2 ging die Kritik hervor, dass bisherige Untersuchungen zur Algorithmen-Aversion wesentliche Faktoren des Arbeitsumfelds von Forecastern nicht berücksichtigen. Es wurde in Studie 1 gezeigt, dass es durchaus solche Faktoren gibt, die sich auf die Algorithmen-Aversion auswirken und somit von praktischer wie auch theoretischer Relevanz sind. Diese Relevanz von Faktoren aus dem Arbeitsumfeld wird im Folgenden weiter vertieft.

Im realen Arbeitsumfeld befinden sich Forecaster mit ihrer Arbeitsleistung bzw. ihren Forecasts nicht in einer mit dem Algorithmus isolierten Entscheidungssituation. In ihrem Arbeitsumfeld interagieren sie mit ihren Kollegen und Konkurrenten – ihren Peers. Der Vergleich mit den Peers hat einen maßgeblichen Einfluss auf das Verhalten von Forecastern und sollte daher im Kontext der Algorithmen-Aversion verstanden werden. Dabei wird bereits in der Literatur zur Algorithmen-Aversion darauf hingewiesen, dass Peers einen Einfluss auf das Verhalten von Forecastern haben und dabei zum Beispiel die Einschätzung der eigenen Leistung beeinflussen: „Furthermore, an extensive literature on overconfidence repeatedly demonstrates that individuals routinely report excessive confidence in their own judgment relative to that of their peers“<sup>554</sup>. Dies lässt erahnen, dass der Entscheidungsprozess durch die Zusammenarbeit mit Algorithmen komplexer wird, wenn Forecaster mit ihrer Leistung in Relation zu ihren Peers stehen.

Studie 2 befasst sich daher mit dem Einfluss von Peers auf die Algorithmen-Aversion. Dies fokussiert auf die Leistung der Forecaster relativ zur Leistung der Peers und dem daraus resultierenden Wettbewerb. Wettbewerbe sind fest in Gesellschaft und Unternehmen verankert und gelten als Erfolgsantrieb.<sup>555</sup> Für diesen Antrieb nehmen Kenntnisse über den Rang und die relative Performance in einem Wettbewerb eine wesentliche Rolle ein. Unternehmen machen sich diesen Mechanismus zu Nutze und stellen ihren Mitarbeitern, vermehrt relatives Performancefeedback zur Verfügung. Hierdurch sollen die Mitarbeiter zu einer höheren Leistung motiviert werden – ohne dabei auf monetäre Anreize zurückgreifen zu müssen.<sup>556</sup>

---

<sup>554</sup> Logg et al. 2019, S. 91.

<sup>555</sup> Vgl. Woike und Hafenbrädl 2020, S. 523.

<sup>556</sup> Die Grundlagen zur Algorithmen-Aversion, sowie die zur Einordnung der Algorithmen-Aversion in das Controlling sind hier gleichermaßen von Relevanz wie in Studie 1.

## Studie 2

Aufbauend auf die Motivation zu Studie 2 in Abschnitt 5.1, werden in Abschnitt 5.2 die theoretischen Rahmenbedingungen für die Einführung von RPI beschrieben. Darauf aufbauend werden die Hypothesen zur Wirkung von RPI auf die Algorithmen-Aversion hergeleitet. Analog zu Studie 1 erfolgen in Abschnitt 5.3 und 5.4 das Experiment und die Hypothesentests mit Ergebnisdiskussion. Daran schließen in Abschnitt 5.4 die Schlussfolgerungen und die Limitationen von Studie 2 an.

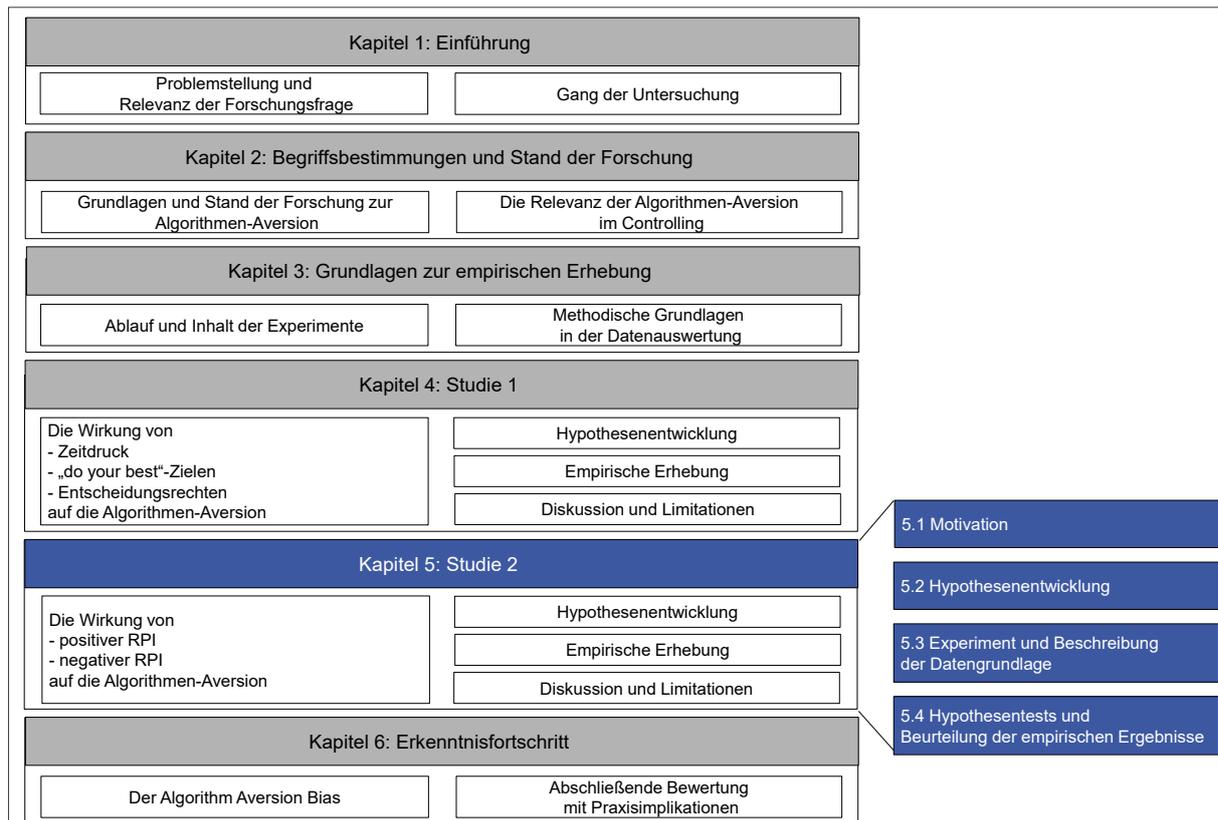


Abbildung 43: Aufbau von Kapitel 5<sup>557</sup>

### 5.1 Motivation

„[R]ank among our equals, is, perhaps, the strongest of all our desires, and our anxiety to obtain the advantages of fortune is accordingly much more excited and irritated by this desire, than by that of supplying all the necessities and conveniences of the body, which are always very easily supplied“<sup>558</sup>.

Die Kenntnis zum relativen eigenen Rang gegenüber Peers, führt zu Wettbewerben innerhalb von Peer-Groups die die Leistung aller Beteiligten steigern.<sup>559</sup> Den daraus erwachsenden Selbstzweck von Informationen zu Rängen beschreiben Blanes i Vidal

<sup>557</sup> Eigene Darstellung.

<sup>558</sup> Smith 1759, S. 169.

<sup>559</sup> Vgl. Woike und Hafenbrädl 2020, S. 523.

und Nossol (2011) mit: „[I]t may be that receiving feedback on relative performance has in itself substantial consequences for employee satisfaction, motivation, and productivity“<sup>560</sup>. Dabei kann ein Rang als ein konstitutives Merkmal eines kompetitiven Umfeldes gesehen werden.<sup>561</sup> Wettbewerbe, führen dazu, dass sich Individuen nicht nur auf ihre eigene, absolute Leistung konzentrieren, sondern ein Fokus auf den Rang innerhalb dieses Wettbewerbs erzeugt wird.<sup>562</sup>

Die Bereitstellung von Feedback, über die relative Leistung eines Feedbackempfängers im Vergleich zu einer Peer-Group, wird in der Forschung auch als Relative Performance Information (RPI) bezeichnet.<sup>563</sup> Die *Performance* stellt im Kontext dieser Arbeit weiterhin ein Synonym für die Forecastinggenauigkeit dar (siehe Abschnitt 2.1.4.2). RPI ist ein nützliches Managementinstrument, denn es kann dazu dienen, durch gesteigerte Motivation ein höheres Leistungsniveau zu erreichen. Es ist daher Teil vieler Controlling- und Anreizsysteme.<sup>564</sup> Meist wird die Erhöhung der Motivation damit begründet, dass RPI soziale Vergleiche ermöglicht und hierdurch Mitarbeitern bewusst wird, wie zufriedenstellend sie ihre Arbeit im Vergleich zu anderen Mitarbeitern erledigen.<sup>565</sup> Im Idealfall kann durch RPI eine ausgewogene Wettbewerbskultur geschaffen werden.<sup>566</sup> RPI wird dabei häufig losgelöst von der individuellen Incentivierung gegeben. Das heißt, die Mitarbeiter erfahren ihren relativen Rang zu ihren Kollegen, werden jedoch nach ihrer individuellen absoluten Leistung entlohnt.<sup>567</sup>

In der unternehmerischen Praxis werden Peer-Group-Vergleiche immer bedeutender. Begünstigt wird die Verbreitung von RPI durch stetig flacher werdende organisationale Strukturen. Flachere Strukturen erhöhen das Potenzial für Peer-Groups in der Leistungsbeurteilung bzw. dem Leistungsfeedback.<sup>568</sup> Einen Katalysator für RPI stellt die digitale Transformation von Unternehmen dar. Durch die elektronische Datenerfassung, -speicherung und -auswertung sind dem Feedbackumfang und der

---

<sup>560</sup> Blanes i Vidal und Nossol 2011, S. 1721.

<sup>561</sup> Vgl. Gjedrem und Kvaløy 2020, o. S.

<sup>562</sup> Vgl. Tran und Zeckhauser 2012, S. 645.

<sup>563</sup> Vgl. Hannan et al. 2013, S. 553; Hannan et al. 2008, S. 894; Holderness et al. 2020, S. 137.

<sup>564</sup> Vgl. Berger et al. 2019, S. 20.

<sup>565</sup> Vgl. Holderness et al. 2019, S. 1202.

<sup>566</sup> Vgl. Woike und Hafenbrädl 2020, S. 523.

<sup>567</sup> Vgl. Song et al. 2018, S. 2628. Siehe hierzu auch das Praxisbeispiel eines virtuellen Callcenters in Silverman 2011.

<sup>568</sup> Vgl. Stewart et al. 2012, S. 435; Wirthmann 2014.

## Studie 2

Feedbackfrequenz kaum mehr Grenzen gesetzt.<sup>569</sup> Dieser Anstieg individueller Daten kann wiederum zu Peer-Group-Daten aggregiert und damit für Leistungsvergleiche genutzt werden.<sup>570</sup> Unternehmen sind dadurch vermehrt imstande, RPI schnell zu erfassen und effektiv bereitzustellen.<sup>571</sup>

In der Praxis findet RPI mittlerweile ein breites Anwendungsspektrum: Es werden unter anderem Kosten optimiert, indem Werksleiter über die Kosten anderer Werke informiert werden, der Umsatz im Vertrieb wird erhöht, indem Vertriebsmitarbeitern die Umsätze anderer Vertriebsmitarbeiter demonstriert werden, oder Ergebnisse werden optimiert, indem Geschäftsführer von Geschäftseinheiten über die Ergebnisse anderer Einheiten informiert werden.<sup>572</sup> Holderness et al. (2017) sprechen davon, dass bereits über 90 % der Fortune 1000 RPI einsetzen.<sup>573</sup>

In der Praxis hat sich jedoch gezeigt, dass RPI bzw. Leistungsvergleiche eine negative Begleitfolge haben.<sup>574</sup> Eine Rankingmethode, die zu negativen Effekten führt, ist die *Rank and Yank*-Methode, oder auch *Vitalitätskurve* genannt, welche durch den ehemaligen General-Electric-Geschäftsführer Jack Welch bekannt wurde.<sup>575</sup> Unter Rank and Yank wird ein Performancevergleich verstanden, bei dem Vorgesetzte ihre Mitarbeiter gemäß deren Leistung einem Rang bzw. einer Rangposition zuordnen. Die Ränge werden anschließend in ein 20-70-10-System aufgeteilt.<sup>576</sup> Dabei gilt: Wer den oberen 20 % zugeordnet wird, zählt zu den besten und leistungsstarken Mitarbeitern. Besonders bedeutsam ist demgegenüber die Zuordnung zu den unteren 10 %. Sie werden als leistungsschwache Mitarbeiter klassifiziert und meist entlassen, um sie durch bessere Mitarbeiter zu ersetzen.<sup>577</sup> Dies übt einen hohen Druck auf die Mitarbeiter aus, wodurch Team-Gedanken in den Hintergrund rücken können und das Selbstvertrauen verloren gehen kann.<sup>578</sup> Namhafte Unternehmen wie Microsoft oder Yahoo machten in der Vergangenheit von der Methodik Gebrauch.<sup>579</sup> Dabei führte

---

<sup>569</sup> Vgl. Alder und Ambrose 2005, S. 45.

<sup>570</sup> Vgl. Alder 2007, S. 157.

<sup>571</sup> Vgl. Holderness et al. 2017, S. 37.

<sup>572</sup> Vgl. Kramer et al. 2016, S. 16.

<sup>573</sup> Gestützt wird diese Aussage durch Daten von Rose und Justine 2013.

<sup>574</sup> Vgl. Holzhacker 2019, S. 51.

<sup>575</sup> Vgl. Faragher 2021, o. S.

<sup>576</sup> Vgl. Holzhacker 2019, S. 51.

<sup>577</sup> Vgl. Pepitone 2013, o. S.

<sup>578</sup> Vgl. Fong et al. 2019, S. 122.

<sup>579</sup> Vgl. Holzhacker 2019, S. 50.

Microsofts ehemaliges Ranking-System zu einer kontraproduktiven Unternehmenskultur, in der nach geraumer Zeit nicht mehr die Leistungsverbesserung im Vordergrund stand.<sup>580</sup> Der Fokus der Mitarbeiter rückte, von Unsicherheit getrieben, häufig weg von effektiver Leistungssteigerung hin zu Tätigkeiten, die die Führungskräfte beeindrucken.<sup>581</sup> Weitere Kritik wird darin gesehen, dass RPI die absolute Leistung vernachlässigt. Dabei kann es durchaus eintreten, dass alle Zielvorgaben von einem Mitarbeiter erreicht werden, dieser jedoch in einem sehr starken Team nur als unterdurchschnittlich eingestuft wird.<sup>582</sup>

Zusammengefasst lässt sich festhalten, dass RPI eine wachsende Relevanz für Unternehmen hat. Darüber hinaus zeigt die Praxis jedoch auch, dass es zu unerwünschten Nebenwirkungen kommen kann. Es ist daher von hoher Relevanz zu verstehen, wie sich RPI auf Aufgaben auswirkt, bei denen eine Algorithmen-Aversion vorliegt. Burton et al. (2019) haben theoretische Überlegungen zur Wirkung von RPI und einer Algorithmen-Aversion angestellt:

„For example, if a decision maker is incentivized to make the best decision (relative to peers on a case-by-case competition basis) rather than a good decision (relative to one's own performance in the long run), then he or she would need to find a way to gain a unique advantage over competitors. If all competitors have access to the same or similar algorithmic aids, then the decision maker would put him or herself at a disadvantage by utilizing the algorithmic judgement because this would mean simply mirroring, rather than surpassing, the performance of other decision makers“<sup>583</sup>.

Diese Überlegungen scheinen stark vereinfacht und lassen einige entscheidende Aspekte außer Acht. Burton et al. (2019) nehmen eine völlig rationale Entscheidungsfindung bei Forecastern an und berücksichtigen nicht, dass jeder Forecaster eine Algorithmen-Aversion bei seinen Peers antizipieren müsste. Das heißt, sie gehen davon aus, dass jeder Forecaster den Algorithmus wählt und sich der Einzelne nur besser stellen kann, wenn er ihn nicht wählt. Das hieße jedoch, dass die Forecaster im Kollektiv davon ausgehen, dass sie besser als der Algorithmus und gegebenenfalls ihre Peers sind. Dies legt nahe, dass unter RPI eine vollständige Algorithmen-Aversion vorherrschen müsste.

---

<sup>580</sup> Vgl. O'Conner 2021, o. S.

<sup>581</sup> Vgl. Holzhacker 2019, S. 51.

<sup>582</sup> Vgl. O'Conner 2021, o. S.

<sup>583</sup> Burton et al. 2019, S. 5.

Wenn Forecaster jedoch durch RPI eine Algorithmen-Aversion bei ihren Peers antizipieren, sollte durch das RPI ein tiefgreifender Denkprozess angestoßen werden, der in der Konsequenz zu einem reflektierten Entscheidungsprozess hinsichtlich der Beurteilung der eigenen Leistung und der des Algorithmus führt. Diese Reflexion kann der besseren Einordnung der eigenen Leistung dienen, obwohl die Incentivierung nicht abhängig von der Rangposition, sondern von der individuellen Leistung ist. Unter dieser Prämisse liegt der Fokus von RPI auf der besseren Reflexion der eigenen Leistung durch eine dritte Dimension des Leistungsfeedbacks (RPI). Aufbauend auf diesen Überlegungen wird im folgenden Abschnitt gezeigt, wie die Wirkung von RPI in der Algorithmen-Aversion aus theoretischer Sicht erwartet zu erwarten ist.

### 5.2 Hypothesenentwicklung

Es wird im Folgenden die Wirkung von RPI auf Feedbackempfänger diskutiert. Im Anschluss werden zwei Theorien vorgestellt, die diese Wirkung teilweise erklären können. Dies ist einerseits die Bereitstellung von Feedback in der Feedback Intervention Theory (FIT) und andererseits der soziale Vergleich in der Social Comparison Theory (SCT). Die beiden Theorien liefern in Kombination einen hohen Erklärungsgehalt und werden in der Literatur ebenfalls häufig gemeinsam betrachtet.<sup>584</sup> Sie bilden anschließend die Basis zur Herleitung der Hypothesen.

Eine Vorbemerkung sei an dieser Stelle, dass in experimentellen Untersuchungen das Leistungsniveau häufig losgelöst von der tatsächlichen Leistung kommuniziert wird.<sup>585</sup> Holderness et al. (2017) konstatieren dazu: „In many studies, negative feedback is completely unrelated to the actions of the participant [...]“<sup>586</sup>. Auch in dieser Studie wird die RPI den Teilnehmern losgelöst von deren tatsächlicher Leistung gegeben. Für positive RPI entstehen daraus keine Implikationen, da positives Feedback in der Regel bereitwillig angenommen wird. Für negative RPI muss jedoch in Teilnehmer mit guter und solche mit schlechter Leistung unterschieden werden.<sup>587</sup> Denn die Wirkung von negativem Feedback kann sich nur entfalten, wenn es nicht zur Verwerfung

---

<sup>584</sup> Siehe zum Beispiel Blanes i Vidal und Nossol 2011.

<sup>585</sup> Vgl. Podsakoff und Farh 1989, S. 45; Vancouver und Tischner 2004, S. 1092.

<sup>586</sup> Holderness et al. 2017, S. 38.

<sup>587</sup> Siehe hierzu beispielsweise auch Pyszczynski et al. 1985, S. 195: „[F]ailure subjects requested more information when they expected it to reveal that most students performed poorly than when they expected it to reveal that most students performed well; success subjects showed little interest in this additional information, regardless of their expectancies as to what it would reveal“.

des Feedbacks durch fehlenden Glauben an die Richtigkeit des Feedbacks kommt.<sup>588</sup> Durch die leistungsunabhängige Bereitstellung ergibt sich damit ein zusätzlicher Erkenntnisgewinn für die Wirkung negativer RPI auf Forecaster mit unterschiedlichem Leistungsniveau.

### 5.2.1 Vertiefung der theoretischen Grundlagen zu RPI

Die empirischen Evidenzen zur Wirkung von RPI basieren zumeist auf Beobachtungs- und Experimentalstudien. Blanes i Vidal und Nossol (2011) betrachten beispielsweise Personalakten in einem Unternehmen, das Akkordlohn bezahlt.<sup>589</sup> Sie heben hervor, dass es durch die Einführung von RPI zu einer ausgeprägten und langanhaltenden Steigerung der Produktivität kommt – ohne, dass das Unternehmen in Form von monetären Anreizen investieren musste. Im Bildungskontext führen Azmat und Iriberry (2010) ein natürliches Experiment durch.<sup>590</sup> Die Schüler einer High School erhalten RPI im Klassenverbund. RPI ist hier die Information darüber, ob ein Schüler über oder unter dem Klassendurchschnitt liegt und wie weit er von diesem Durchschnitt entfernt ist. Sie kommen zu dem Ergebnis, dass durch die Einführung von RPI eine Verbesserung der Leistung um 5 % eintritt.

Kramer et al. (2016) weisen in einem Laborexperiment mit 175 Studenten einer europäischen Universität einen positiven Zusammenhang zwischen der Bereitstellung von RPI und der Steigerung der Performance nach.<sup>591</sup> Tiefergehendes Ziel der Arbeit von Kramer et al. (2016) ist die Untersuchung der Wirkung des Detailgrads und der Darstellungsform von RPI auf die Leistung. Neben der Verbesserung der Performance werden jedoch keine Effekte entdeckt. Das heißt, dass sich die Darstellung eines spezifischen Rangs in seiner Wirkung nicht verändert, wenn er von „schlecht zu gut“, „gut zu schlecht“ oder randomisiert dargestellt wird.<sup>592</sup>

Falk und Ichino (2006) untersuchen Peer-Feedback, indem sie die Teilnehmer einer Studie Briefe in Umschläge stecken ließen.<sup>593</sup> Es wird dabei eine feste, nicht leistungsabhängige Vergütung festgelegt. Dabei waren die Teilnehmer entweder zu zweit in einem Zimmer (RPI möglich) oder allein (keine RPI möglich). Sie erkennen bei

---

<sup>588</sup> Vgl. Carroll und Marginson 2021, o. S.; Falk und Ichino 2006, S. 39.

<sup>589</sup> Vgl. Blanes i Vidal und Nossol 2011, S. 1730.

<sup>590</sup> Vgl. Azmat und Iriberry 2010, S. 435.

<sup>591</sup> Vgl. Kramer et al. 2016, S. 16.

<sup>592</sup> Vgl. Kramer et al. 2016, S. 21.

<sup>593</sup> Vgl. Falk und Ichino 2006, S. 39.

## Studie 2

Teilnehmern, die zu zweit in einem Raum sind, eine starke Verbesserung der Produktivität bei geringerer Standardabweichung. Besonders hervorgehoben wird, dass Teilnehmer bei schlechter Leistung<sup>594</sup> eine besondere Sensitivität für RPI aufweisen.<sup>595</sup>

Neben der Betrachtung von positiver Wirkung durch RPI gibt es Studien, die negative Effekte zeigen. Guryan et al. (2009) untersuchen wie Falk und Ichino (2006) Peer-Effekte in Zweiertteams.<sup>596</sup> Dabei werfen sie die Frage auf: „Is an employee’s productivity influenced by the productivity of his or her nearby co-workers?“<sup>597</sup> Sie versuchen dies anhand von empirischen Daten aus professionellen Golfspielen abzuleiten. In einem Golfspiel werden zwei Spieler unterschiedlichen Spielniveaus zufällig einander zugewiesen, woraus sie auf die Präsenz von RPI schließen. Sie finden dabei keine Hinweise darauf, dass die Leistung (*Performance*) eines Golfers von den Fähigkeiten (*Ability*) seines zugewiesenen Mitspielers beeinflusst wird.

Hannan et al. (2008) zeigen Unterschiede in der Wirkung von RPI, die von der zugrunde gelegten Art der Incentivierung abhängen.<sup>598</sup> Für eine Incentivierung, die lediglich an die eigene Leistung gekoppelt ist, finden sie eine positive Wirkung auf die Leistung durch RPI – unabhängig von der Darstellung des Feedbacks. Wenn jedoch ein sehr präzises Feedback gegeben wird und ein *Tournament Incentive*<sup>599</sup> zugrunde liegt, verschlechtert RPI die Leistung der Feedbackempfänger. Hier muss angemerkt werden, dass in der Tournament-Incentivierung nur die besten 10 % der Teilnehmer der Studie einen Bonus bekamen.<sup>600</sup> Auch, wenn dies durch einen hohen Bonus für diese 10 % zum gleichen Erwartungswert zwischen den Gruppen führte, ist nicht ausgeschlossen, dass Effekte von RPI hier als Forecast für die Aussicht auf den Bonus interpretiert wurden. RPI kann bei einer Tournament-Incentivierung also wie ein Forecast für die zu erwartende Vergütung wirken und kann damit auch ein Indikator für eine unerreichbare Incentivierung sein.<sup>601</sup> Ein Leistungsabfall ist damit erklärbar.

---

<sup>594</sup> Falk und Ichino 2006, S. 39 sprechen im Original von „low-productivity workers“.

<sup>595</sup> Vgl. Falk und Ichino 2006, S. 39.

<sup>596</sup> Vgl. Guryan et al. 2009, S. 34.

<sup>597</sup> Guryan et al. 2009, S. 34.

<sup>598</sup> Vgl. Hannan et al. 2008, S. 893.

<sup>599</sup> Bei einem *Tournament Incentive* wird die Höhe der Vergütung in Abhängigkeit der relativen Leistung gegenüber der von Kollegen gesetzt. Siehe Hannan et al. 2008, S. 902.

<sup>600</sup> Vgl. Hannan et al. 2008, S. 902.

<sup>601</sup> Vgl. Tafkov 2013, S. 329.

Diese kurze Übersicht einiger Studien zu RPI zeigt, dass RPI einen Einfluss auf die Leistung hat. Dieser Einfluss ist zumeist positiv und der Einsatz von RPI verbessert die Leistung. Einige Studien zeigen jedoch einen negativen Zusammenhang von RPI auf die Leistung. Im Folgenden werden zwei Theorien vorgestellt, die die Wirkung von RPI tiefergehend erklären und die Basis für die Ableitung von Hypothesen zur Wirkung von RPI, im Kontext der Algorithmen-Aversion, darstellen.

### 5.2.1.1 Feedback Intervention Theory

Kluger und DeNisi (1996) sehen grundsätzlich eine differenzierte Wirkung von Feedback auf die Leistung.<sup>602</sup> In einer Metastudie mit 131 Studien betonen sie zwar, dass Leistungsfeedback im Schnitt die Leistung von Feedbackempfängern verbessert, darüber hinaus gibt es jedoch einen beträchtlichen Anteil von einem Drittel an Studien, die durch Feedback eine negative Wirkung bzw. eine Verschlechterung der Leistung bescheinigen. Dabei ist es zunächst nicht relevant, ob es sich um positives oder negatives Feedback handelt – für beide Fälle finden sich positive wie auch negative Auswirkungen auf die Leistung. Da es demnach keine einfachen Annahmen und weitreichenden Erklärungen dazu gab, wie Feedback wirkt, entwarfen Kluger und DeNisi (1996) ein mittlerweile weit verbreitetes Erklärungsmodell zur Wirkung von Feedback – die FIT.<sup>603</sup> Die FIT soll Erklärungsansätze für die unterschiedliche Wirkungsweise von Feedback liefern.

Die FIT ist in der Controlling-Literatur weit verbreitet, da durch diese das Verhalten von Forecastern erklärt und die optimale Steuerung untersucht werden kann.<sup>604</sup> Dabei steht die Leistung häufig im Mittelpunkt. Besonders wertvoll wird die Erkenntnis über die Differenzierung verschiedener Ebenen, in denen Feedback wirkt, gesehen: „To predict how individuals will respond to negative performance feedback, we draw upon Feedback Intervention Theory, which states that feedback effectiveness is dependent on whether the recipient of the feedback focuses his or her attention on the task level or the self-level“<sup>605</sup>. Um ein Verständnis für die Ebenen des Task Level und des Self-Level zu entwickeln, bedarf es fünf zentraler Annahmen in der FIT:<sup>606</sup>

---

<sup>602</sup> Siehe auch Abschnitt 2.1.4.2.

<sup>603</sup> Vgl. DeNisi und Kluger 2000; Kluger und DeNisi 1996.

<sup>604</sup> Siehe beispielsweise Hannan et al. 2008; Holderness et al. 2017; Lourenco et al. 2018; Newman et al. 2021.

<sup>605</sup> Holderness et al. 2017, S. 34.

<sup>606</sup> Adaptiert von Kluger und DeNisi 1996, S. 259.

## Studie 2

- Das Verhalten wird durch den Leistungsvergleich von Feedback mit Zielen oder Anforderungen beeinflusst.
- Ziele oder Anforderungen können hierarchisch organisiert werden.
- Die Aufnahmefähigkeit des Feedbackempfängers ist begrenzt. Nur die Leistungsvergleiche, die Aufmerksamkeit erhalten, können das Verhalten beeinflussen.
- Aufmerksamkeit ist in der Regel auf eine moderate Hierarchieebene gerichtet.
- Leistungsvergleiche können den Fokus der Aufmerksamkeit verändern und damit das Verhalten beeinflussen.

Die FIT erklärt die Wirkung von Feedback schlussendlich über die Wirkebene, auf der das jeweilige Feedback einen Feedbackempfänger anspricht. Das heißt, viel bedeutsamer als der Fakt, dass Feedback negativ oder positiv ist, ist die Ebene, die das Feedback beim Empfänger anspricht. Kluger und DeNisi (1996) unterscheiden dies in drei hierarchische Wirkebenen: „[*Meta-task processes* involving the self, *task-motivation processes* involving the focal task, and *task-learning processes* involving the task details of the focal task [Hervorhebungen im Original]“<sup>607</sup>. Im Folgenden werden diese Ebenen bezeichnet als *Metaebene*, *Motivationsebene* und *Lernebene*, wobei die Motivations- und die Lernebene auch als *Aufgabenebene* zusammengefasst werden können. Die Meta- und die Motivationsebene vereint, dass auf diesen Ebenen Feedback einen wertenden Charakter besitzt. Ohne diesen wertenden Charakter und mit direktem Bezug zur Aufgabe und etwaigen Handlungsempfehlungen steht die Lernebene. Dies macht deutlich, dass die Ebenen in der Realität nicht trennscharf anzutreffen sind.

Feedback in der Motivations- und der Lernebene stellen zum Beispiel in einer Präsentation die Rückmeldung über die korrekte Länge der Präsentation oder konkrete inhaltlich Verbesserungsvorschläge dar. Der Vortragende kann sich unmittelbar an einer zeitlichen Abweichung orientieren und die Präsentation kürzen. Im Gegensatz dazu handelt es sich um ein Feedback auf der Metaebene, wenn zum Beispiel die Sprechfertigkeit des Vortragenden thematisiert wird. Diese kann im Sinne einer Belohnung und Bestrafung gesehen werden, die positiv wie negativ wirken und den Fokus von der eigentlichen Aufgabe weglenken kann. Kluger und DeNisi (1998) betonen,

---

<sup>607</sup> Kluger und DeNisi 1996, S. 262.

## Studie 2

dass Feedback auf der Metaebene ein höheres Risiko hat, nicht leistungsfördernd oder gar leistungshemmend zu wirken.<sup>608</sup>

Feedback, welches konstruktiv auf den Aufgabenebenen gegeben wird, führt zu einer Steigerung der zukünftigen Leistung in zumeist der bestehenden Arbeitsweise. Das Feedback wird zu Informations- und Lernzwecken genutzt und dient unmittelbar der Motivation. Lernprozesse auf der Lernebene werden insbesondere durch negatives Feedback angestoßen. Durch das negative Feedback erhöhen Feedbackempfänger ihre Anstrengungen und versuchen, die Aufgabe besser zu erfüllen, indem sie ihr Verhalten entsprechend dem Erlernten anpassen. Dies geschieht in der Regel zunächst durch einen intensiveren Arbeitseinsatz (*working harder*). Falls dies nicht direkt gelingt, wird auf der Lernebene versucht, eine neue Strategie zu finden, um die Aufgabe zu erfüllen (*working smarter*).<sup>609</sup>

Feedback, das auf der Motivationsebene beim Empfänger ankommt, ist eng an die zugrunde liegende Aufgabe gekoppelt und auf einem gemäßigten positiven oder negativen Niveau. Es wird in der Regel eine Rückmeldung zum Leistungsstand im Vergleich zu einem im Vorfeld definierten Ziel gegeben. Ist diese Rückmeldung realistisch und positiv, wird das Anstrengungsniveau in der Regel beibehalten. Ist die Rückmeldung realistisch und negativ, motiviert sie dazu, die Leistung zu steigern.<sup>610</sup> Wenn das Feedback deutlichere Abweichungen hat, können zwei Effekte auftreten: Einerseits kann eine Motivation zum Lernen eintreten, bei der versucht wird, aus bisherigen Arbeitsschritten zu lernen, um künftig besser abzuschneiden. Andererseits kann eine Verschiebung der Aufgabenerfüllung zu neuen Strategien erfolgen oder die Aufmerksamkeit wird hin zur Metaebene und dem Selbstbild stattfinden.<sup>611</sup>

Dies wiederum kann zur Verwerfung des Feedbacks führen, um das Selbstbild zu schützen. Feedback auf der Metaebene kann die Ziele der Aufgabe auf eine andere Ebene lenken oder das Ziel gänzlich anders ausrichten: „Meta-task processes include processes that link the focal task with higher order goals, such as the evaluation of the implication of task performance for the self“<sup>612</sup>. Feedback, das auf der Metaebene emp-

---

<sup>608</sup> Vgl. Kluger und DeNisi 1998, S. 69.

<sup>609</sup> Vgl. Kluger und DeNisi 1996, S. 263.

<sup>610</sup> Vgl. Kluger und DeNisi 1996, S. 261.

<sup>611</sup> Vgl. Kluger und DeNisi 1996, S. 261.

<sup>612</sup> Vgl. Kluger und DeNisi 1996, S. 262.

## Studie 2

fangen wird, führt dadurch häufig zu einer Senkung der zukünftigen Leistung. Feedbackempfänger nutzen ihre kognitiven Ressourcen, um die potenzielle Bedrohung des Selbstbilds durch das Feedback abzuwenden, und vernachlässigen die erforderlichen Maßnahmen, um in der Aufgabe besser zu werden.<sup>613</sup> Feedback auf einer Metaebene kann entstehen, wenn der Feedbackempfänger ein stark negatives oder stark positives Feedback zur Aufgabe selbst erhält, das ihn persönlich betrifft – also sein Selbstbild –, oder wenn das Feedback auf eine normative Ebene – wie einen Vergleich zum sozialen Umfeld – abzielt.<sup>614</sup>

Da selbst ein Lob zu einer geringeren Leistung führen kann, wenn die Aufmerksamkeit weg von der Aufgabe und hin zum Selbstbild gerichtet wird, nehmen die Wirkebenen folglich eine zentrale Rolle im Verständnis von Feedback ein. Feedback aus dem sozialen Umfeld kann dabei beim Empfänger den Selbstwert senken, den Fokus auf neue Aufgaben legen und eine Abkehr von der zu erfüllenden Aufgabe bedeuten, wodurch Motivation und Leistung sinken.<sup>615</sup> Weiter ist dazu festzuhalten, dass die hierarchisch höheren Ebenen die darunterliegenden überlagern können. Auch, wenn die Wirkebenen durch die Art, wie Feedback gegeben wird, wechseln können, sind Feedbackempfänger weitestgehend starr auf eine Wirkebene fokussiert. Das heißt, wenn das Feedback einmal auf eine bestimmte Ebene gelenkt ist, bleibt es im Wesentlichen auf diese Ebene begrenzt. Hierzu ist außerdem anzumerken, dass es bei Feedbackempfängern eine Tendenz zur Mitte, also auf die Motivationsebene, gibt. Es gibt darüber hinaus eine generelle Aversion gegen Feedback auf der Metaebene, die sehr stark auf das Selbstbild abzielt. Gleichzeitig sind die Lerneffekte auf der Lernebene teilweise abgedeckt durch Automatismen und erfordern keine tiefere Aufmerksamkeit.<sup>616</sup>

Ergänzend zu diesen Ausführungen werden die Effekte der Algorithmen-Aversion betrachtet. Das Feedback in der Forschung zur Algorithmen-Aversion berücksichtigt bislang die Leistung des Algorithmus und die eigene Leistung. Es wäre denkbar, dass diese Art des Feedbacks die Teilnehmer anregt, in der Lernebene zu verweilen, ergo einem working-harder-Prinzip zu folgen. Um in einem intensiveren Wettbewerb

---

<sup>613</sup> Vgl. Holderness et al. 2017, S. 34.

<sup>614</sup> Vgl. Holderness et al. 2017, S. 34.

<sup>615</sup> Vgl. Kluger und DeNisi 1996, S. 262.

<sup>616</sup> Vgl. Kluger und DeNisi 1996, S. 262. Siehe hierzu auch Carver und Scheier 1981; Wicklund 1975.

effektiv zu agieren, müssen Forecaster ihre Arbeitsstrategien und Arbeitsweisen jedoch anpassen.<sup>617</sup> Das heißt, neben den gewohnten Arbeitsweisen müssen sie neue Möglichkeiten suchen, wie sie das Ziel eines effektiven Forecasts erreichen, und damit verstärkt den Algorithmus berücksichtigen.<sup>618</sup> Durch RPI kann die Ebene, auf der Feedback wirkt, verändert werden. Ein Wechsel der Feedbackebene kann verschiedene Mechanismen erzeugen. Vom Übergang der Aufgaben-Lern-Ebene in die Metaebene werden Emotionen beim Feedbackempfänger geweckt. Die Verarbeitung und Einordnung dieser Emotionen verlangt nach kognitivem Aufwand und überlagert womöglich Lern- und Motivationseffekte. Die Theorie legt hier zwar zunächst nahe, dass es durch Feedback zu Peers zu einer Verschiebung auf die Metaebene kommt, jedoch gibt es dazu wiederum eine Vielzahl an Hemmnissen, wie die Tendenz zur Mitte, also zur Motivationsebene, und die grundsätzliche Abneigung beim Feedbackempfänger gegen Metaebenenfeedback. Daraus ergibt sich eine Wirkung von RPI auf das Verhalten im Forecasting und der Algorithmen-Aversion, die sich aufbauend auf den Überlegungen zur FIT zeigt. Diese Wirkung kann als Informationseffekt verstanden werden. Das heißt, die Information zur relativen Leistung wird herangezogen, um die Information zur absoluten Leistung zu vergleichen und damit besser einzuordnen. Dadurch kann die Entscheidung aus der absoluten Leistung verifiziert werden. Dies führt einerseits zu einer Erleichterung in Entscheidungsprozessen und andererseits hilft es dabei, die eigene Leistung besser zu reflektieren.<sup>619</sup>

### 5.2.1.2 Social Comparison Theory

Die positive Wirkung von RPI wird unmittelbar deutlich in *Tournament-Kontrakten*.<sup>620</sup> *Tournaments* können als Wettbewerb verstanden werden, wobei die Vergütung, also der Tournament-Kontrakt, eine spezielle Form der leistungsabhängigen Vergütung im Wettbewerb darstellt. Hier entsteht durch RPI eine Prognose zur Leistung gegenüber Peers: „In a tournament contract, RPI informs the agent of the likelihood of winning the tournament reward and, thus, of the marginal benefit of additional effort“<sup>621</sup>. Das heißt, in einem Tournament wird ein Forecaster anhand seiner relativen Position zu seinen

---

<sup>617</sup> Vgl. Chen und Keith 2005, S. 11; Magretta 2002, S. 90.

<sup>618</sup> Vgl. Webb et al. 2013, S. 1434.

<sup>619</sup> Vgl. Newman et al. 2021, S. 75.

<sup>620</sup> Siehe auch Abschnitt 5.2.1.

<sup>621</sup> Tafkov 2013, S. 329.

## Studie 2

Peers entlohnt. RPI steht demnach in direkter Korrelation zur Vergütung. Eine Verbesserung der relativen Rangposition geht mit einer Verbesserung der Entlohnung einher, wodurch der Effekt auf die Motivation und Leistung unmittelbar erklärbar ist.

RPI hat darüber hinaus jedoch auch eine Wirksamkeit, wenn diese direkte Korrelation zwischen relativer Leistung bzw. Rangposition und Incentivierung nicht vorliegt. Unter der Annahme rationaler Entscheidungsfindung dürfte RPI jedoch keine Wirkung haben, wenn die Information zur relativen Leistung keine Konsequenz für die Vergütung hat. In der herkömmlichen Incentivierung, wie sie auch in dieser Arbeit vorgenommen wird, ist der Erwartungswert der Vergütung unabhängig von der Leistung Anderer und eine relative Verbesserung der Leistung wird nicht besser entlohnt.<sup>622</sup> Für eine Incentivierung, die allein von der individuellen Leistung abhängt und damit keine Abhängigkeit der individuellen Vergütung von der Leistung bzw. Vergütung Anderer aufweist, tritt dennoch ein Einfluss auf den Feedbackempfänger auf. Dies begründet sich darin, dass durch das Feedback über die erreichte und auch die nicht erreichte Rangposition der Drang zu sozialen Vergleichen entsteht.<sup>623</sup>

Ein prominenter Effekt von RPI wird durch die SCT erklärt, denn RPI ermöglicht immer einen sozialen Vergleich mit Peers, der sich nachweislich auf die Leistung des Feedbackempfängers auswirkt.<sup>624</sup> Die SCT besagt im Kern, dass Forecaster einen natürlichen Drang danach haben, sich mit anderen zu vergleichen, um ihr Selbstbild oder auch ihre Leistung und Fähigkeiten zu bewerten.<sup>625</sup> Der Terminus *social comparison* wurde zum ersten Mal von Festinger (1954) verwendet. Seitdem wird die SCT von unterschiedlichen Fachrichtungen wie Psychologie oder Sozialwissenschaften angewendet und weiterentwickelt.<sup>626</sup>

Ein sozialer Vergleich kann zur Reflektion und Einschätzung der eigenen Fähigkeiten herangezogen werden<sup>627</sup>, das Selbstbild verändern<sup>628</sup> oder die Stimmungslage bzw. Emotionen<sup>629</sup> und die Motivation<sup>630</sup> beeinflussen. Die Beeinflussung der

---

<sup>622</sup> Vgl. Tafkov 2013, S. 329.

<sup>623</sup> Vgl. Woike und Hafenbrädl 2020, S. 526.

<sup>624</sup> Vgl. Holderness et al. 2020, S. 138.

<sup>625</sup> Vgl. Festinger 1954, S. 117.

<sup>626</sup> Für einen historischen Abriss zur SCT siehe Suls und Wheeler 2000.

<sup>627</sup> Vgl. Mussweiler et al. 2004, S. 833.

<sup>628</sup> Vgl. Morse und Gergen 1970, S. 148.

<sup>629</sup> Vgl. Crusius und Mussweiler 2012, S. 780.

<sup>630</sup> Vgl. Crusius und Mussweiler 2012, S. 780.

Selbsteinschätzung wirkt umso stärker, wenn objektive Maßstäbe fehlen. Beim Vorliegen objektiver Maße können die Einflüsse des sozialen Vergleichs jedoch ebenfalls auftreten und die objektiven Maße dabei teilweise überlagern.<sup>631</sup> Konsistent mit der FIT sehen Holderness et al. (2020) für die Wirkung von RPI durch die SCT: „[S]ocial comparisons inherent to RPI feedback can divert cognitive resources from on-task activities (e.g., task-related learning), to offtask activities (e.g., worry or anxiety) and/or self-regulatory activities (e.g., processing feedback)“<sup>632</sup>. Durch diese umfassende Anwendungsmöglichkeit und den hohen Erklärungsgehalt findet die SCT in der Controlling-Forschung breite Verwendung.<sup>633</sup> Die Stärke des Effekts durch RPI auf die Leistung lässt sich in der SCT durch drei Kriterien beschreiben:

- Es ist zu berücksichtigen, dass die Motivation zur Leistungsverbesserung nicht für jeden sozialen Vergleich übereinstimmend ist. Die Motivation zur Leistungsverbesserung wird insbesondere aus dem Vergleich mit Peers gezogen. Peers sind Menschen, die in gewisser Weise Ähnlichkeit zum Feedbackempfänger aufweisen. Die Wahrnehmung einer Ähnlichkeit und damit die Entstehung eines Peereffekts erfordert Ähnlichkeiten in Attributen, die die Leistung determinieren.<sup>634</sup> Dies kann beispielsweise durch ähnliche Anstrengung (*effort*), ähnliche Fähigkeiten (*ability*) oder, ganz konkret, ähnliches Gehalt entstehen.<sup>635</sup>
- Neben den Ähnlichkeiten mit Peers ist es relevant für die Wirksamkeit eines Peer-Vergleichs, dass eine gleiche oder vergleichbare Aufgabe innerhalb der Peer-Group erbracht wird.<sup>636</sup> Sind die zu erbringenden Aufgaben zu verschieden, besitzt der betrachtete Peer keine Relevanz für die Evaluation der eigenen Leistung in der zugrunde liegenden Aufgabe.
- Es ist bedeutsam, dass der Vergleich in einem für den RPI-Empfänger relevanten Bereich erfolgt. Je wichtiger dieser Bereich ist, desto größer ist der innere Drang, sich durch RPI selbst zu beurteilen.<sup>637</sup> Der relevante Bereich kann beispielsweise ebenfalls das Gehalt oder auch persönliche Merkmale betreffen.

---

<sup>631</sup> Vgl. Klein et al. 1990, S. 179.

<sup>632</sup> Holderness et al. 2020, S. 138.

<sup>633</sup> Vgl. Berger et al. 2019; Carroll und Marginson 2021; Hannan et al. 2013; Holderness et al. 2017; Knauer et al. 2017; Kramer et al. 2016; Mahlendorf et al. 2014; Tafkov 2013.

<sup>634</sup> Tafkov 2013, S. 330 spricht hier von „characteristics that predict performance“.

<sup>635</sup> Vgl. Cornelissen et al. 2017, S. 427.

<sup>636</sup> Vgl. Tafkov 2013, S. 330.

<sup>637</sup> Vgl. Tafkov 2013, S. 330.

## Studie 2

Je stärker diese drei Kriterien erfüllt sind, desto stärker kann durch RPI das Selbstbild beeinflusst werden. Die SCT besagt hier, dass durch sozialen Vergleich, das Selbstbild geprägt wird und dabei jeder Mensch nach einem positiven Selbstbild strebt. Im Umkehrschluss wird versucht, negative Gefühle über das Selbstbild zu vermeiden und ein positives Selbstbild zu erhalten. Dies führt dazu, dass Forecaster sich in dem, was sie bereits gut können, noch mehr anstrengen und ein positive RPI erhalten, da sie mit ihrer guten Leistung ein positives Selbstbild bestätigen. Ergo wirkt eine positive RPI bestärkend auf die Art und Weise, wie eine Leistung erbracht wird, da diese das Selbstbild stärkt. Demgegenüber werden für negative RPI Strategien für das Verhalten ersucht, die einen Ausweg aus den negativen Gefühlen ermöglichen.<sup>638</sup>

Eine bedeutende Nebenbedingung zur positiven Wirkung von RPI wird durch die *Theory of Downward Comparison* als Teilbereich der SCT formuliert.<sup>639</sup> Menschen streben grundsätzlich nach einem positiven Selbstbild. Dieses Selbstbild kann sich jedoch durch negative Erfahrungen wie unzureichende Leistungen im Beruf verschlechtern. Durch Peervergleiche mit Peers, die schlechter abschneiden, kann wiederum ein aufhellendes, positives Selbstbild zur eigenen Leistung entstehen, wenn eine Überlegenheit zu diesen Peers wahrgenommen wird.<sup>640</sup> Es ist demnach weniger relevant, wie gut ein Mensch seine individuelle Leistung erbringt, denn, wenn durch RPI eine Abwärtsvergleich möglich ist, wird sich seine Art und Weise, die Leistung zu erbringen, bestärken.

Die SCT liefert damit eine Erklärung, weshalb RPI auch bei individueller Incentivierung einen Effekt hat und wann dieser besonders ausgeprägt ist. Je stärker das Selbstbild beeinflusst wird, desto stärker fällt der Effekt aus. Es ist hierzu festzuhalten, dass nicht jedes Feedback durch RPI unmittelbar beim Feedbackempfänger eine völlige Torsion des Selbstbilds auslöst. Aus den Ausführungen zur FIT ist bekannt, dass RPI mehrere Ebenen beim Feedbackempfänger gleichzeitig ansprechen kann.<sup>641</sup> Hinsichtlich der Algorithmen-Aversion in Experimentalsettings ist daher die Wirkung von RPI vor dem Hintergrund der nicht vorhandenen persönlichen Beziehungen der Probanden untereinander eher auf einem gemäßigten Niveau angesiedelt. Weiter liegen in den Experimentalsettings objektive Maßstäbe vor, die eine starke Wirkung auf das

---

<sup>638</sup> Vgl. Tafkov 2013, S. 329–330.

<sup>639</sup> Vgl. Gibbons et al. 2002; Wills 1981.

<sup>640</sup> Vgl. Alicke und Govorun 2005, S. 85.

<sup>641</sup> Siehe Abschnitt 5.2.1.1.

Selbstbild verhindern. Das heißt, RPI regt durchaus zum Verifizieren der eigenen Leistung an, wird jedoch nicht das Selbstbild stark verändern können. Darüber hinaus trägt die SCT zur adäquaten Gestaltung des Experiments bei, denn damit RPI wirken kann, müssen die drei oben genannten Kriterien erfüllt sein.

### 5.2.1.3 Zusammenfassung der theoretischen Grundlagen

Die Erkenntnisse der potenziellen Wirkung von RPI durch die FIT und die SCT als Verbindungsglied zur Algorithmen-Aversion werfen die Frage auf, wie sich die Einführung von RPI auf die Algorithmen-Aversion auswirkt. Sowohl FIT als auch SCT prognostizieren, dass durch RPI eine weitere Dimension in der Wirkung des Leistungsfeedbacks Einfluss auf das Verhalten von Forecastern nimmt. Wie dieser Einfluss erwartet wird, wird in der folgenden Hypothesenentwicklung dargestellt.

In der Literatur zur Algorithmen-Aversion gibt es bislang keine Untersuchung, die sich mit dem Thema RPI auseinandersetzt. Bis dato ist Feedback die Gegenüberstellung der Forecasts der Teilnehmer mit dem des Algorithmus – also das vielfach erwähnte Leistungsfeedback. Dieses Leistungsfeedback ist objektiv, dennoch sind die Teilnehmer nicht in der Lage, das Feedback richtig einzuordnen. Sie überschätzen ihren eigenen Forecast.<sup>642</sup> Zum Messen der Algorithmen-Aversion wird in der Literatur teilweise ein WOA untersucht, bei dem die Gewichtung eines des Ratschlags von entweder einem Algorithmus oder einem anderen Forecaster gewertet wird.<sup>643</sup> Dies steht jedoch nicht im Zusammenhang mit dem Peer-Feedback, da der Ratschlag zur Unterstützung des eigenen Forecasts herangezogen werden kann und nicht zur Beurteilung der eigenen Leistung dient. In der realen Umgebung wird die Leistung in der Forecastinggenauigkeit jedoch nicht nur absolut zum Forecaster oder relativ zum Algorithmus betrachtet. Jeder Forecaster steht auch in Interaktion und Wettbewerb zu seinen Peers und kann damit eine RPI erhalten.

Wenn Feedback relativ formuliert wird, gibt es immer positives wie auch negatives Feedback. Aus der SCT ist ersichtlich, dass für die Wirkung von RPI daher auch in positive und negative RPI unterschieden werden muss. Positive RPI bestärkt den Feedbackempfänger in dem was er tut, negative RPI lässt ihn verzagen oder wenn

---

<sup>642</sup> Vgl. Dietvorst et al. 2018, S. 1161, 2015, S. 120.

<sup>643</sup> Siehe beispielsweise Logg et al. 2019, S. 96.

möglich, neue Lösungsstrategien ersuchen. Es wird in der folgenden Hypothesenentwicklung daher zwischen positiver und negativer RPI unterschieden.<sup>644</sup>

### 5.2.2 Der Einfluss von positiver RPI auf die Algorithmen-Aversion

Im Allgemeinen zeigen Feedbackempfänger eine Präferenz für positives Feedback gegenüber negativem Feedback auf, da Ersteres ein positives Stimmungsbild fördert. Anders als negatives stellt positives Feedback in der Regel das Selbstbild nicht in Frage und wird bereitwillig akzeptiert und als zutreffend wahrgenommen.<sup>645</sup> Positives Feedback führt zu einer gelassenen Betrachtung einer Aufgabe und gleichzeitig zu einer starken Überzeugung der eigenen Leistung.<sup>646</sup> Dies wiederum führt zur Überzeugung, auch in Zukunft – im Falle der vorliegenden Untersuchung: dem incentivierten Forecast – eine gute Leistung zu erbringen.

Für positive RPI gibt es ausgehend von der SCT eine eindeutige Tendenz der erwarteten Wirkung. Forecaster, die im Rahmen von RPI ein positives Peer-Feedback erhalten, finden Bestätigung für ihre Leistung. Die Bestätigung und Überzeugung zur eigenen Leistung führen zu einem verstärkten Fokus auf die eigene Leistung. Der relevante Lernprozess zur reflektierten Beurteilung der eigenen Leistung und der des Algorithmus wird damit durch RPI weiter vernachlässigt. Die Bestätigung der eigenen Leistung wird demnach zu einer Zunahme der Wahl des eigenen Forecasts führen.

Unterstützung findet diese Argumentation in der FIT, denn Feedbackempfänger nehmen auch hier positives Feedback bereitwillig an. Ein solches, vom Feedbackempfänger als wahr und realistisch angesehenes Feedback, ist eine positive Verstärkung der eigenen Leistung und führt zu einer Bestärkung der Art der Leistungserbringung (*working harder*) und damit zu sinkendem Algorithmeinsatz.

In diesen Überlegungen darf nicht vernachlässigt werden, dass in Forecasts mit Algorithmen-Aversion außerdem ein objektives absolutes Leistungsfeedback gegeben wird. Objektive Kriterien mildern die Effekte, die auf das Selbstbild wirken, ab. Dies spricht gegen eine Veränderung der Wirkebene auf die Metaebene oder eine starke Veränderung der Wahrnehmung des eigenen Forecasts. Nichtsdestotrotz liegt eine zusätzliche Informationsfunktion von RPI im Falle von positivem Leistungsfeedback

---

<sup>644</sup> Vgl. Tafkov 2013, S. 330.

<sup>645</sup> Vgl. Holderness et al. 2017, S. 35.

<sup>646</sup> Vgl. Kuhnen und Tymula 2012, S. 94.

vor und wird die Forecaster in der Wahl zu ihrem eigenen Forecast bestätigen. Daraus ergibt sich die Hypothese H5.<sup>647</sup>

H5: Positive RPI zur Forecastinggenauigkeit erhöht die Algorithmen-Aversion.

### 5.2.3 Der Einfluss von negativer RPI auf die Algorithmen-Aversion

Negatives Feedback durch RPI kann unterschiedliche Auswirkungen auf den Feedbackempfänger haben. Es kann dazu führen, dass Forecaster ihre Leistung steigern, um einen höheren relativen Rang zu erreichen.<sup>648</sup> Aus negativem Feedback resultiert jedoch auch ein steigender Druck auf die individuelle Leistung.<sup>649</sup> Dieses Verhalten wurde bereits in Abschnitt 4.2.2 aufgegriffen und als *choking under pressure* beschrieben.<sup>650</sup> Dabei wird argumentiert, dass sich die Leistung von Forecastern verschlechtern kann, wenn sie aufgrund von Druck bei einer Aufgabe nicht mehr die gewünschte Leistung erbringen können.<sup>651</sup> Das führt dazu, dass Forecaster mit schlechter Leistung sich durch die RPI demoralisieren lassen und ihre Leistung reduzieren.<sup>652</sup> Darüber hinaus kann eine Art Sorge um das Selbstbild entstehen. Alder (2007) merkt hier an: „Some cues may direct attention away from the task toward other goals of the self (meta-task processes) and consequently hinder performance“<sup>653</sup>. Trotz der potenziell negativen Wirkung von RPI auf das Selbstbild ist die Darstellung von negativem Feedback im Kontext von RPI unvermeidbar.<sup>654</sup>

Um negativer RPI entgegenzuwirken, sucht die menschliche Psyche nach Ausweichstrategien.<sup>655</sup> Das heißt, es werden alternative Erklärungen für das schlechte Abschneiden gesucht, die sich aus externen Umständen ergeben und sich nicht auf die eigenen Fähigkeiten beziehen.<sup>656</sup> Dieses Verhalten wird allgemein auch als *Self-Serving Bias* beschrieben.<sup>657</sup> Das heißt, nach negativer RPI sinkt die Bereitschaft, an einer

---

<sup>647</sup> Um eine ganzheitliche Diskussion der Ergebnisse von Studie 1 und Studie 2 zu vereinfachen, werden die Hypothesen studienübergreifend durchnummeriert.

<sup>648</sup> Vgl. Tafkov 2013, S. 330.

<sup>649</sup> Vgl. McCarty 1986, S. 840.

<sup>650</sup> Vgl. Beilock und Carr 2001, S. 701.

<sup>651</sup> Vgl. Dai et al. 2018, S. 1667.

<sup>652</sup> Vgl. Ashraf et al. 2014, S. 44; Bandiera et al. 2013, S. 1079.

<sup>653</sup> Alder 2007, S. 163.

<sup>654</sup> Vgl. Fong et al. 2019, S. 121. Fong et al. (2019) definieren negatives Feedback als eine negative Bewertung, die von einem Individuum über ein Produkt, eine Leistung oder über die Charakteristika eines anderen Individuums abgegeben werden, wobei der Feedbackgeber die Bewertungsgrundlage festlegt.

<sup>655</sup> Vgl. Arnold et al. 2018, S. 1.

<sup>656</sup> Vgl. Bénabou und Tirole 2002, S. 871; Kuhnen und Tymula 2012, S. 96.

<sup>657</sup> Vgl. Bénabou und Tirole 2002, S. 871.

## Studie 2

Aufgabe teilzunehmen, wenn die Chancen zur Leistungsverbesserung in der Zukunft als schlecht eingestuft werden.<sup>658</sup> Diese Reaktion stellt im Kontext der Interaktion mit Algorithmen einen interessanten Aspekt dar. Wenn die RPI als eine Kritik an der aktuellen Leistung interpretiert wird, kann der Algorithmus als Ausweichstrategie fungieren. Durch die Wahl auf den Algorithmus finden Forecaster einen Ausweg aus der Kritik und können sie von sich selbst wegleiten. Die künftige Teilnahme an der Aufgabe wird durch den Algorithmus erfüllt.

Aus der FIT und der SCT ist bekannt, dass Feedback unterschiedlich stark auf den Feedbackempfänger wirkt. Bei einer sehr starken Bindung zu Peers und einem persönlich formulierten Feedback wird das Selbstbild bzw. die Metaebene angesprochen. Diese tiefgreifende Wirkung kann jedoch im vorliegenden Experiment zur Algorithmen-Aversion nicht angenommen werden. Es wird zwar durch RPI das Selbstbild angesprochen, jedoch wird es aufgrund der geringen Bindung zu den Peers und der Arbeitsumgebung zu keiner Verwerfung des Selbstbilds kommen. Neben der Wirkung auf das Selbstbild ist insbesondere die Informationsfunktion von RPI daher von Relevanz. Durch die Einführung von negativer RPI wird den Teilnehmern der Studien aus einer zweiten Quelle ihre schlechte Leistung bewusst gemacht. Dies ermöglicht eine bessere Reflektion der eigenen Leistung, wodurch sich ein verstärkter Einsatz von Algorithmen begründen würde. RPI kann demnach aus Sicht der SCT sowohl als soziale Motivation als auch als Informationsfunktion wirken und dabei jeweils die Algorithmen-Aversion senken. Die FIT legt nahe, dass durch die negative RPI und damit das zweifach negative Feedback zur Leistung eine tendenzielle Abkehr von der Lernebene beim Empfänger stattfindet. Dies hat zur Folge, dass die Intensivierung des Arbeitseinsatzes (*working harder*) nicht weiterverfolgt wird und auf der Lern- und Motivations-ebene versucht wird, eine neue Strategie zu finden, um die Aufgabe zu erfüllen (*working smarter*).<sup>659</sup> Dies würde ebenfalls dem Self-Serving Bias folgen.

Die Wirkung von negativer RPI wird über die nähere Betrachtung des Leistungsniveaus des Feedbackempfängers weiter differenziert.<sup>660</sup> Falk und Ichino (2006) zeigen, dass Feedback an sich eine geringere Wirkung auf die Leistung von leistungsstarken wie auf die von leistungsschwächeren Forecastern hat. FIT und SCT können

---

<sup>658</sup> Vgl. Köszegi 2006, S. 673.

<sup>659</sup> Vgl. Kluger und DeNisi 1996, S. 263.

<sup>660</sup> Vgl. Delfgaauw et al. 2014, S. 2.

## Studie 2

dies durch die geringe Veränderung der Wirkebene erklären.<sup>661</sup> Leistungsstärkere Forecaster sind überzeugter von ihrer Leistung und dadurch weniger empfänglich dafür, Feedback auf einer höheren Wirkebene einzuordnen. Insbesondere bei schlechter Performance wirkt RPI hingegen besonders stark. Hier decken sich vor allem das Empfinden und die tatsächliche Leistung bei schlechter Leistung mit negativer RPI und es findet eine Bestätigung der schlechten Leistung beim Feedbackempfänger statt.<sup>662</sup> Falls sich das Empfinden der eigenen Leistung nicht mit der RPI deckt, kann dies zur Verwerfung der RPI führen.<sup>663</sup> Für das Vorliegen von negativer RPI wird daher vorhergesagt, dass nur die Algorithmen-Aversion der Forecaster mit schlechter Leistung beeinflusst wird. Es wird dazu die Hypothese H6 aufgestellt:

H6: Negative RPI zur Forecastinggenauigkeit senkt bei leistungsschwächeren Forecastern die Algorithmen-Aversion.

Für Teilnehmer mit guter Leistung, die negatives Peer-Feedback erhalten, wird keine Hypothese aufgestellt. Es wird erwartet, dass ihr Verhalten sich nicht wesentlich verändert.

### 5.2.4 Zusammenfassung der Hypothesen

Tabelle 25 fasst die erwarteten gegenläufigen Effekte in den Hypothesen in Studie 2 zusammen. Im Folgenden wird beschrieben, welche notwendigen Änderungen im Experimentaldesign zur Untersuchung dieser Hypothesen vorgenommen werden und wie die RPI operationalisiert wird.

Tabelle 25: Zusammenfassung der Hypothesen in Studie 2

Hypothese	Prognostizierte Wirkung
H5: Positive RPI zur Forecastinggenauigkeit erhöht die Algorithmen-Aversion.	↑
H6: Negative RPI zur Forecastinggenauigkeit reduziert bei leistungsschwächeren Forecastern die Algorithmen-Aversion.	↓

<sup>661</sup> Vgl. Falk und Ichino 2006, S. 36.

<sup>662</sup> Vgl. Falk und Ichino 2006, S. 39.

<sup>663</sup> Vgl. Carroll und Marginson 2021, o. S.

## 5.3 Experiment und Beschreibung der Datengrundlage

### 5.3.1 Experimentaldesign und Operationalisierung der Manipulation

Grundsätzlich ist die Untersuchung in Studie 2 an das Vorgehen in Abschnitt 3.2 und Abschnitt 4.3 angelehnt. Das heißt, die im Folgenden als Kontrollgruppe bezeichnete Gruppe entspricht der Algorithmen-Aversions-Kontrollgruppe (AAC). Die Algorithmen-Aversion wurde bereits in Studie 1 gezeigt (H1), weshalb sie in Studie 2 nicht nochmals untersucht wird. Das heißt, für Studie 2 erhalten alle Teilnehmer Leistungsfeedback.

Die wesentliche Veränderung im Experimentaldesign ist die Einführung von RPI als Treatment. Insgesamt werden die Teilnehmer der Treatmentbedingungen dreimal auf die RPI aufmerksam gemacht. Zunächst vor Beginn der Übungsforecasts, nach fünf Übungsforecasts und vor dem incentivierten Forecast. Vor Beginn der zehn Übungsforecasts wurde den Teilnehmern, die RPI erhalten, der Text in Abbildung 44 gezeigt.

We have already conducted this survey a year ago. To help you assessing your performance, we will provide the performance data of other MTurkers to you. To evaluate your progress we provide you with this assessment after the first 5 practice estimates and after all 10 practice estimates.  
Your performance can either be in the best, middle oder worst third compared to other MTurkers performance.

Next

Abbildung 44: Erster Hinweis zur RPI<sup>664</sup>

Im Falle von positiver RPI haben die Teilnehmer nach fünf Übungsforecasts zusätzlich den Text in Abbildung 45 gezeigt bekommen. Der untere Teil aus Abbildung 45 wurde für negative RPI angepasst (siehe Abbildung 46). Es wird dabei das Peer-Feedback eingeführt (worst/best 33 %). Hierbei wurde Podsakoff und Farh (1989) folgend eine eindeutig positive bzw. negative RPI festgelegt.<sup>665</sup> Diese wurde, ebenfalls den Autoren folgend, durch einen Pretest auf Wirkung und Glaubwürdigkeit untersucht. Dieser Pretest wurde am 16.02.2021 mit zehn Personen am International Performance Research Institute in Stuttgart durchgeführt. Dabei unterlief jeder Teilnehmer des Pretests mehrere Runden im Experiment mit unterschiedlicher RPI. Dies reichte von

<sup>664</sup> Eigene Darstellung (Auszug aus dem Experiment).

<sup>665</sup> Vgl. Podsakoff und Farh 1989, S. 49.

## Studie 2

worst/best 40 % hin zu worst/best 20 %. Anschließend wurde in einer Gruppendiskussion das Empfinden für das jeweilige Feedback diskutiert. Dabei erwiesen sich worst/best 33 % als glaubhaft und einschneidend beim Feedbackempfänger.

Performance evaluation on your first 5 practice estimates:

We also conducted this survey a year ago. At that time, we recorded the needed time and the performance of MTurkers **incentivized estimate**.

**For their incentivized estimate, the other MTurkers could choose to have either their own estimated rank or the statistical model's estimated rank determine their bonus. In other words, they chose to be paid based on their accuracy, or chose to be paid based on the model's accuracy.**

We also measured your needed time and performance in the first 5 practice estimates.

Compared to the other MTurkers with a similar amount of time needed for their incentivized forecast, your accuracy is within the

top	33 %
-----	------

This means, that you performed

better than at least	66 %
----------------------	------

of the MTurkers who have done this survey before.

Next

Abbildung 45: Treatment für positive RPI<sup>666</sup>

Compared to the other MTurkers with a similar amount of time needed for their incentivized forecast, your accuracy is within the

worst	33 %
-------	------

This means, that you performed

worse than at least	66 %
---------------------	------

of the MTurkers who have done this survey before.

Next

Abbildung 46: Treatment für negative RPI<sup>667</sup>

Nach den zehn Übungsforecasts wurde den Teilnehmern nochmals die RPI gezeigt und eine Einschätzung von ihnen verlangt, wie sie die Algorithmen-Aversion bei ihren Peers beurteilen (siehe Abbildung 47 und Tabelle 3). Durch diese zusätzliche Frage zur Interaktion mit dem Algorithmus sind die Teilnehmer gezwungen, sich kognitiv mit ihrer Peer-Group auseinanderzusetzen. Dies erhöht die Aufmerksamkeit für

<sup>666</sup> Eigene Darstellung (Auszug aus dem Experiment).

<sup>667</sup> Eigene Darstellung (Auszug aus dem Experiment).

## Studie 2

die Peer-Group und damit die Wirksamkeit für RPI. Ferner kann durch diese Frage eine Indikation zur Wirksamkeit des Treatments erhoben werden.

Performance evaluation on your 10 practice estimates:

**For their incentivized estimate, the other MTurkers could choose to have either their own estimated rank or the statistical model's estimated rank determine their bonus. In other words, they chose to be paid based on their accuracy, or chose to be paid based on the model's accuracy.**

We also measured your needed time and performance within all 10 practice estimates.

Compared to the other MTurkers with a similar amount of time needed for their incentivized forecast, your accuracy now is within the

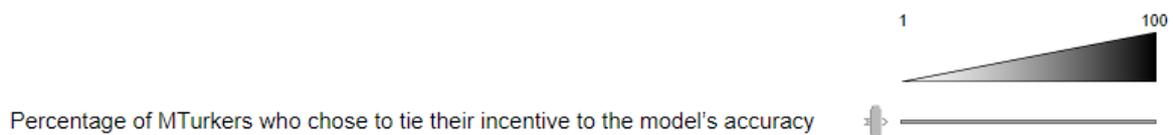
top	33 %
-----	------

This means, that you performed

better than at least	66 %
----------------------	------

of the MTurkers who have done this survey before.

**What do you think, how many of the other MTurkers chose to tie their incentive to the model's accuracy?**



Next

Abbildung 47: Finale Darstellung der RPI<sup>668</sup>

Hannan et al. (2008) folgend, wurde den Teilnehmern mitgeteilt, dass die Aufgabe die sie zu bearbeiten haben, bereits einmal durchgeführt wurde und sich das Peer-Feedback an der Leistung dieser Teilnehmer orientiert.<sup>669</sup> Dies ist jedoch nur eine grobe Erklärung der Peer-Group. Damit RPI wirken kann, benötigen die Teilnehmer eine direkte Verbindung zu ihren Peers, also eine Gruppe von Teilnehmern, die ihnen ähnlich ist.<sup>670</sup> Ein solches Ähnlichkeitsmerkmal sollte messbar sein und ex ante abgeglichen werden können. Das heißt, es muss ein Merkmal sein, das in der Aufgabe begründet ist und nicht in den persönlichen Merkmalen. Diese Verbindung wird durch einen ähnlichen *Effort* erzeugt.<sup>671</sup> Für die zu Grunde liegende Forecastingaufgabe wird Efendić et al. (2020) gemäß auf den Zeitaufwand je Teilnehmer für den Forecast verwiesen.<sup>672</sup> Die für eine Aufgabe benötigte Zeit spiegelt die Schwierigkeit und den damit

<sup>668</sup> Eigene Darstellung (Auszug aus dem Experiment).

<sup>669</sup> Vgl. Hannan et al. 2008, S. 902.

<sup>670</sup> Vgl. Carroll und Marginson 2021, o. S.; Cornelissen et al. 2017, S. 427 und Abschnitt 5.2.1.2.

<sup>671</sup> Siehe Abschnitt 5.2.1.2.

<sup>672</sup> Vgl. Efendić et al. 2020, S. 104.

## Studie 2

einhergehenden bzw. getätigten Arbeitsaufwand wider.<sup>673</sup> Es kann suggeriert werden, dass schnelle Teilnehmer mit schnellen Teilnehmern verglichen werden und langsamere Teilnehmer mit langsamen. Den Teilnehmern wird vor Beginn der Trainingsrunden dieses Vorgehen mitgeteilt und eine hochzählende Uhr als Referenz für den Effort angezeigt.

Die Erhebung in Studie 2 erfolgte zweistufig. Die Nomenklatur der finalen Experimentalgruppen findet sich in Tabelle 26. Um die Kontrollgruppe nicht mehrfach zu erheben und damit ressourcenschonend in der Datenerhebung vorzugehen, wurde die Erhebung zweistufig durchgeführt (siehe Abbildung 48). Zunächst wurden die Teilnehmer für H5, also im Falle von positiver RPI, jeweils einer Kontroll- und eine Treatmentgruppe (positive RPI) zugewiesen. Ein Gruppenvergleich ist anschließend möglich.

Tabelle 26: Nomenklatur der Experimentalgruppen in Studie 2

<b>Bezeichnung</b>	<b>Leistungsfeedback (Treatment)</b>	<b>Performance</b>
<b>CA</b>	Kontrollgruppe (C)	Alle Teilnehmer (A)
<b>PA</b>	Positive RPI (P)	Alle Teilnehmer (A)
<b>CGP</b>	Kontrollgruppe (C)	Gute Performance (GP)
<b>CBP</b>	Kontrollgruppe (C)	Schlechte Performance (BP)
<b>NGP</b>	Negative RPI (N)	Gute Performance (GP)
<b>NBP</b>	Negative RPI (N)	Schlechte Performance (BP)

Für H6 müssen die Teilnehmer in solche mit guter und solche mit schlechter Leistung unterteilt werden. Das Leistungsniveau der Teilnehmer ist ex ante nicht bekannt und es muss bei der Einteilung berücksichtigt werden, dass durch RPI in der Regel das Leistungsniveau leicht ansteigt.<sup>674</sup> Daher werden die Kontrollgruppe und die Gruppe mit negativer RPI anhand eines Median-Splits mit eingeschlossenen oberen Endpunkten hinsichtlich der Forecastinggenauigkeit eingeteilt.<sup>675</sup> Die Forecastinggenauigkeit wird aus den Übungsrunden ermittelt.<sup>676</sup> So ergeben sich jeweils 50 %

<sup>673</sup> Siehe dazu auch Hannan et al. 2013, S. 559.

<sup>674</sup> Vgl. Podsakoff und Farh 1989, S. 45.

<sup>675</sup> Vgl. Iacobucci et al. 2015, S. 652. Kritik am Median-Split wird durch den Powerverlust gesehen. Die Notwendigkeit zur Aufteilung der Gruppen in gute und schlechte Leistung wurde bereits in der Hypothesenentwicklung und damit vor der Datenerhebung festgelegt. Daher wurden die Gruppengrößen zur Erhebung verdoppelt. Siehe dazu Cohen 1983, S. 249.

<sup>676</sup> Die Teilnehmer anhand in ihrer Leistung bzw. Fähigkeiten in den Übungsforecasts einzuordnen ist adaptiert von Kramer et al. 2016, S. 19. Sie schreiben dazu: „The preliminary round served two purposes. First, it allowed participants to get familiar with the task. Second, and more importantly, it enabled us to get an individual benchmark

## Studie 2

*gute Performer* und 50 % *schlechte Performer* eingeteilt nach Leistungsniveau. Das bedeutet, dass die zu erhebende Gruppengröße bei negativer RPI doppelt so groß sein muss (mindestens 500 Teilnehmer in der endgültigen Stichprobe). Darüber hinaus wird konsistent mit bisherigen Erhebungen zur Algorithmen-Aversion und Studie 1 die Confidenceeinschätzung in den eigenen Forecast und in den des Algorithmus abgefragt (siehe Tabelle 3).

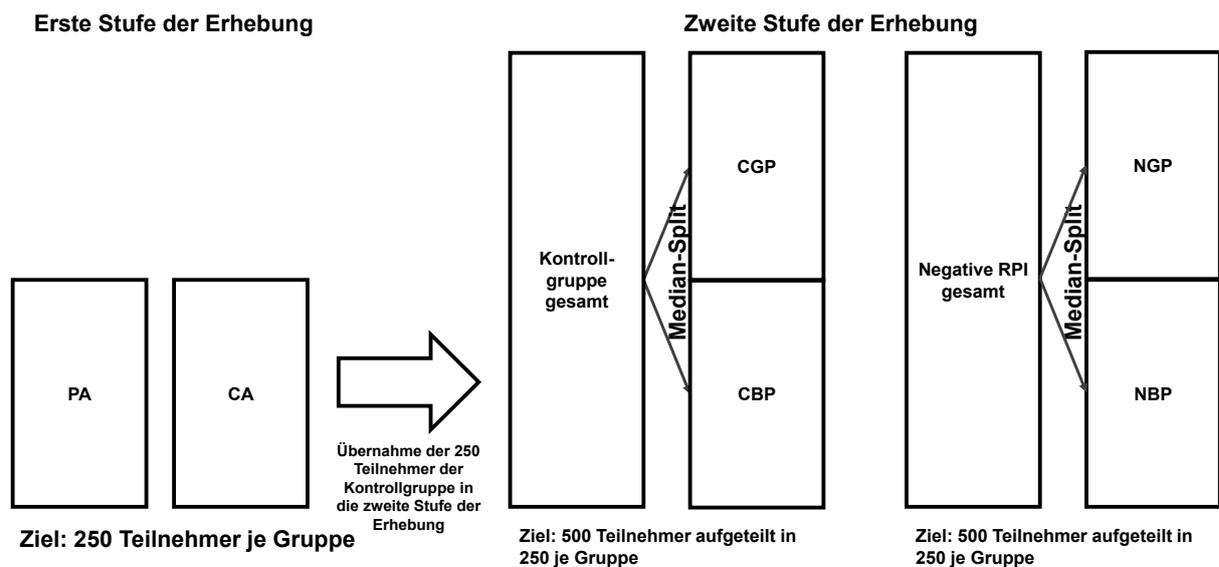


Abbildung 48: Zweistufige Erhebung der Experimentalgruppen in Studie 2<sup>677</sup>

### 5.3.2 Deskriptive Statistik

Die Auswahl und die Bezahlung der Teilnehmer folgen im Wesentlichen dem Vorgehen aus Studie 1. Um eine jeweilige Gruppengröße von mindestens 250 Teilnehmern zu erhalten, werden zunächst 2681 Teilnehmer akquiriert. Aufgrund von Schwankungen in den finalen Gruppengrößen in Studie 1 (siehe Tabelle 9) werden die angestrebten Gruppengrößen leicht überschritten. Die Erhebung erfolgte vom 07.01.2022 bis zum 11.01.2022. Aus dieser Gesamtheit von 2681 werden 519 Teilnehmer ausgeschlossen, da sie die wesentliche abhängige Variable zur Entscheidung zwischen dem eigenen Forecast oder dem des Algorithmus nicht beantwortet haben. Weitere 76 Teilnehmer werden ausgeschlossen, weil sie die Aufmerksamkeitsprüfung und 683, weil sie die Verständnisfragen nicht bestanden haben (siehe dazu Abschnitt 3.2). Somit

performance level for each of the participants. Our hypotheses are about the marginal effects of RPI provision on individual performance“.

<sup>677</sup> Eigene Darstellung.

## Studie 2

ergibt sich eine Stichprobe von 1403 Teilnehmern.<sup>678</sup> In dieser Stichprobe liegt ein ausgeglichenes Verhältnis von Frauen und Männern mit 46 % weiblichen und 54 % männlichen Teilnehmern vor. Die Stichprobe ist im Durchschnitt 39 Jahre alt (SD = 11,66). Dabei sind über 50 % der Teilnehmer unter 36 Jahre alt. Abbildung 49 zeigt die Verteilung des Alters der Stichprobe.

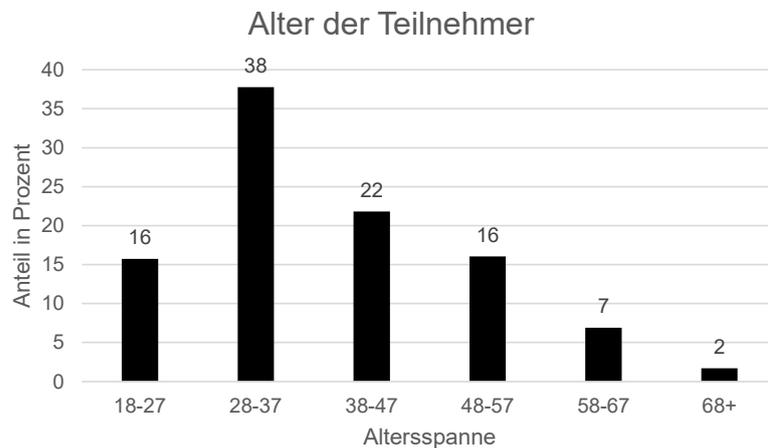


Abbildung 49: Alter der Teilnehmer in Studie 2<sup>679</sup>

Abbildung 50 zeigt die Verteilung des jeweils höchsten Bildungsabschlusses der Teilnehmer in der finalen Stichprobe (ansteigend v. l. n. r.). Mit 54 % hat der Großteil der Teilnehmer ein 4-jähriges College besucht. Dem folgen mit rund 19 % Teilnehmer mit einem Masterabschluss.

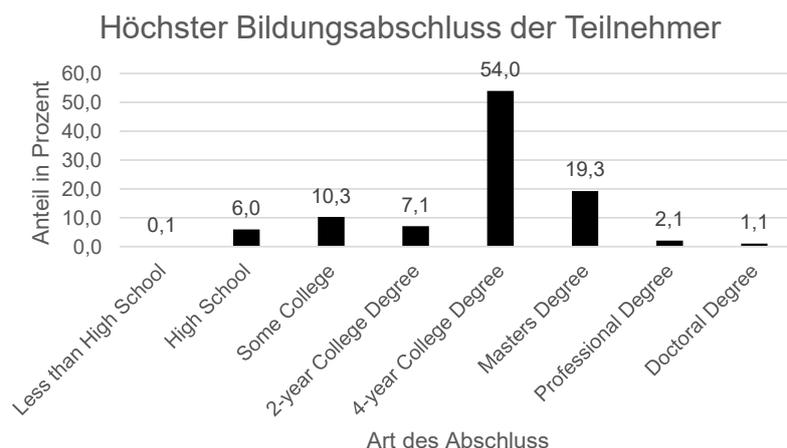


Abbildung 50: Bildungsniveau der Teilnehmer in Studie 2<sup>680</sup>

<sup>678</sup> Dies entspricht eine Ausschlussrate von 47 %.

<sup>679</sup> Eigene Darstellung.

<sup>680</sup> Eigene Darstellung.

## Studie 2

Zu besseren Einordnung und Vergleichbarkeit der Stichproben mit Dietvorst et al. (2015) werden die deskriptiven Werte von Studie 2 denen von Dietvorst et al. (2015) und Studie 1 gegenübergestellt (siehe Tabelle 27). Es zeigt sich, dass die Stichproben der beiden Studien dieser Arbeit grundlegend ähnlich sind. Eine leichte Abweichung ergibt sich beim Alter, denn die Probanden in den beiden Studien dieser Arbeit sind durchweg älter als die Teilnehmer der Studien in Dietvorst et al. (2015).

Tabelle 27: Vergleich der Stichproben Studie 1 und Studie 2 sowie Dietvorst et al. (2015)<sup>681</sup>

	<b>Studie 1</b>	<b>Studie 2</b>	<b>Dietvorst et al. (2015)</b>
<b>Durchschnittsalter</b>	37 – 40	37 – 40	33 – 34
<b>Geschlecht (% weiblich)</b>	43 – 49	43 – 50	46 – 53
<b>Stichprobenumfang (durchschnittliche Gruppengröße)</b>	230	278	259

### 5.3.2.1 Zuteilung der Experimentalgruppen

Alle Teilnehmer in Studie 2 erhalten in zehn Übungsforecasts Leistungsfeedback. Es gibt demnach keine Algorithmen-Appreciation-Gruppe, weshalb eine Unterscheidung zwischen Algorithmen-Aversion und Algorithmen-Appreciation entfällt. Das Treatment ist die positive (P) sowie negative (N) RPI. Ferner wird für H6 zusätzlich unterschieden in gute (GP) und schlechte Leistung (BP) (siehe Tabelle 26).

Studie 2 wird gemäß Abbildung 48 zweistufig durchgeführt, entsprechend werden die jeweils erhobenen Daten chronologisch beschrieben. Für H5 werden Teilnehmer mit positivem RPI mit solchen verglichen, die kein RPI erhalten haben. Das heißt, es wurden hierfür zunächst mindestens 250 Teilnehmer je Gruppe CA und PA erhoben. Die finalen Gruppengrößen finden sich in Tabelle 28.

Tabelle 28: Gruppengrößen zur Untersuchung zu positivem RPI

	<b>CA</b>	<b>PA</b>
<b>Keine Differenzierung der Performance</b>	n = 267	n = 285

Zum Testen der H6 ist es notwendig, die Teilnehmer in solche mit guter und solche mit schlechter Leistung zu unterteilen. Es wird daher die Kontrollgruppe CA

<sup>681</sup> Eine Spanne entsteht durch die Angabe über mehrere Versuchsgruppen hinweg. Siehe Dietvorst et al. 2015, S. 116.

## Studie 2

(siehe Tabelle 28) um weitere 276 Teilnehmer auf insgesamt 543 Teilnehmer erweitert. In der Gruppe mit negativer RPI befinden sich 575 Teilnehmer. Dies ergibt in Summe 1118 Teilnehmer. Von besonderem Interesse für H6 sind die Teilnehmer mit schlechter Performance. Es werden daher jeweils die Teilnehmer mit guter und die mit schlechter Leistung in gute und schlechte Performer mittels Median-Split eingeteilt (siehe Tabelle 29 und Abschnitt 5.3.1). Die Gruppen mit guter Performance sind für die Hypothesentests nicht relevant. Zur Transparenz und Nachvollziehbarkeit der vorliegenden Erhebung, werden sie dennoch in den folgenden Ausführungen diskutiert und in den Tabellen *ausgegraut* dargestellt.

Tabelle 29: Gruppengrößen und Performance der Gruppen zu negativem RPI

	<b>Kontrollgruppe (C)</b>	<b>Negative RPI (N)</b>
<b>Schlechte Performance (BP) bzw. Forecastinggenauigkeit</b>	> 10,1 mit n = 270	>9,3 mit n = 283
<b>Gute Performance (GP) bzw. Forecastinggenauigkeit</b>	<= 10,1 mit n = 273	<= 9,3 mit n = 292

Die randomisierte Verteilung der Stichprobe in die Experimentalgruppen ergibt die finalen Gruppengrößen in Tabelle 30. Die Regel für die Randomisierung lautete *gleichverteilte Ziehung*. Tabelle 30 und Tabelle 31 zeigen die soziodemografischen Daten zwischen den aufgeteilten Experimentalgruppen.

Tabelle 30: Geschlecht und Alter nach Experimentalgruppen in Studie 2

<b>Gruppe</b>	<b>Geschlecht (% weiblich)</b>	<b>Durchschnittsalter</b>
<b>CA (n = 267)</b>	49	38,79
<b>PA (n = 285)</b>	51	39,91
<b>CBP (n = 270)</b>	44	37,83
<b>NBP (n = 283)</b>	46	38,13
<b>CGP (n = 273)</b>	44	39,11
<b>NGP (n = 292)</b>	43	40,92

## Studie 2

Tabelle 31: Bildungsabschluss nach Experimentalgruppen in Studie 2<sup>682</sup>

Gruppe	Bildungsabschluss (Anteil in %)
CA (n = 267)	1 = 0; 2 = 6,7; 3 = 7,9; 4 = 6,4; 5 = 53,9; 6 = 22,5; 7 = 1,1; 8 = 1,5
PA (n = 285)	1 = 0; 2 = 6,7; 3 = 9,1; 4 = 9,1; 5 = 49,8; 6 = 22,1; 7 = 2,8; 8 = 0,4
CBP (n = 270)	1 = 0; 2 = 3,7; 3 = 8,5; 4 = 5,6; 5 = 57; 6 = 23,0; 7 = 1,5; 8 = 0,7
NBP (n = 283)	1 = 0,4; 2 = 4,2; 3 = 6,4; 4 = 4,2; 5 = 62,5; 6 = 20,1; 7 = 1,4; 8 = 0,7
CGP (n = 273)	1 = 0; 2 = 8,1; 3 = 12,8; 4 = 8,8; 5 = 53,8; 6 = 12,8; 7 = 2,2; 8 = 1,5
NGP (n = 292)	1 = 0,3; 2 = 7,2; 3 = 14,4; 4 = 7,9; 5 = 46,9; 6 = 18,5; 7 = 2,4; 8 = 2,4

### 5.3.2.2 Performance und Bezahlung

Die notwendige Vorbedingung, dass eine Algorithmen-Aversion nur als solche gesehen werden kann, wenn die Teilnehmer durch die Wahl des eigenen Forecasts eine schlechtere Performance haben, ist auch für diese Untersuchung gegeben. In engem Zusammenhang damit steht, dass die Teilnehmer durch die Wahl des eigenen Forecasts einen geringen Bonus erhalten. Tabelle 32 erweitert konsistent mit Studie 1 die Performance-Betrachtung zur Forecastinggenauigkeit im incentivierten Forecast und zeigt, dass alle Teilnehmer signifikant schlechter abschneiden als der Algorithmus.

Tabelle 32: Forecastinggenauigkeit von Forecastern und Algorithmus in Studie 2 (M und SD)<sup>683</sup>

Gruppe	Genauigkeit Forecaster	Genauigkeit Algorithmus	Differenz	Paardifferenzentest
CA	9,39 (8,72)	4,49 (4,29)	4,90 (9,24)	t(266) = 8,672, p < 0,001
PA	9,00 (8,92)	4,33 (4,04)	4,67 (9,41)	t(284) = 8,386, p < 0,001
CBP	12,76 (9,79)	4,39 (4,24)	8,38 (10,71)	t(269) = 8,378, p < 0,001
NBP	11,72 (9,73)	4,27 (4,12)	7,46 (10,37)	t(282) = 12,098, p < 0,001
CGP	6,51 (6,34)	4,02 (3,89)	2,49 (6,79)	t(272) = 6,004, p < 0,001

<sup>682</sup> 1 = Less than High School; 2 = High School; 3 = Some College; 4 = 2-year College Degree; 5 = 4-year College Degree; 6 = Masters Degree; 7 = Professional Degree; 8 = Doctoral Degree.

<sup>683</sup> Bei der Interpretation von Tabelle 32 ist zu berücksichtigen, dass die Teilnehmer die außerhalb des Bereichs der Bonuszahlungen lagen dennoch eine Bezahlung i. H. v. 0,01 \$ erhalten haben. Das heißt, alle Forecasts außerhalb des Bonusbereichs, gehen sowohl für den Algorithmus als auch den menschlichen Forecast mit 0,01 \$ ein.

## Studie 2

Gruppe	Genauigkeit Forecaster	Genauigkeit Algorithmus	Differenz	Paardifferenzentest
NGP	7,11 (7,03)	4,49 (4,31)	2,62 (7,64)	t(291) = 5,913, p < 0,001

Tabelle 32 zeigt, dass es über alle Gruppen hinweg einen signifikanten Unterschied in der Forecastinggenauigkeit zwischen dem menschlichen Forecast und dem des Algorithmus gibt. Die Forecasts des Algorithmus schwanken im Mittel zwischen 4,02 und 4,49 und liegen damit im antizipierten Bereich. Die Forecastinggenauigkeit der Forecaster liegt zwischen 6,51 und 12,76. Wie zu erwarten, sind dabei die Gruppen mit guter Leistung auch im incentivierten Forecast besser, als die mit schlechter Leistung.<sup>684</sup> Analog zu Tabelle 32 werden in Tabelle 33 die aus der Leistung resultierenden Boni bzw. die leistungsabhängige Incentivierung dargestellt. Zur Unterstreichung der Relevanz der Algorithmen-Aversion wird abermals der Vergleich zur Vergütung, die durch den Algorithmus möglich ist, aufgezeigt.

*Tabelle 33: Bonuszahlungen in \$ in Studie 2 (M und SD)*

Gruppe	Bonus Forecaster	Bonus Algorithmus	Differenz	Paardifferenzentest
CA	0,28 (0,34)	0,47 (0,38)	-0,19 (0,48)	t(266) = -6,669, p < 0,001
PA	0,31 (0,36)	0,48 (0,36)	-0,17 (0,48)	t(284) = -6,072, p < 0,001
CBP	0,17 (0,29)	0,48 (0,37)	-0,32 (0,47)	t(269) = -11,052, p < 0,001
NBP	0,22 (0,32)	0,49 (0,36)	-0,28 (0,46)	t(282) = -10,062, p < 0,001
CGP	0,37 (0,35)	0,51 (0,37)	-0,14 (0,47)	t(272) = -4,915, p < 0,001
NGP	0,36 (0,36)	0,48 (0,37)	-0,12 (0,46)	t(291) = -4,390, p < 0,001

Die Incentivierung der Teilnehmer greift bis zu einer Abweichung von sechs Rängen in der Forecastinggenauigkeit (siehe Tabelle 2). Durch die verdichtete Leistung der Teilnehmer in den Gruppen mit guter Performance hin zum incentivierten Leistungsbereich bzw. vice versa in Gruppen mit schlechter Performance demonstriert

<sup>684</sup> Siehe dazu Abschnitt 5.2.1.

Tabelle 33, wie verlustreich sich eine Algorithmen-Aversion gestaltet, wenn Forecaster schlechte Leistungen erbringen.

### 5.3.2.3 Algorithmen-Aversion und Confidence

Tabelle 34 zeigt die Confidencewerte. Über alle Gruppen hinweg liegt die Confidence in den Algorithmus signifikant über der in den eigenen Forecast. Die niedrigsten Werte der Confidence in den eigenen Forecast sind dabei in der CGP und der NGP zu finden. Das heißt, bei der besten Leistung (siehe Tabelle 32) haben die Teilnehmer die geringste Confidence in ihren Forecast. Gleichzeitig haben sie auch die geringste Confidence in den Forecast des Algorithmus. Die jeweils höchsten Werte finden sich in der CBP. Bei der schlechtesten Leistung haben die Teilnehmer über alle Gruppen hinweg die höchste Confidence in den eigenen Forecast und in den des Algorithmus. Tabelle 34 zeigt außerdem, dass die Confidence durch RPI flächendeckend sinkt.<sup>685</sup> Dies gilt über alle Gruppen hinweg sowohl für den eigenen Forecast als auch für den des Algorithmus.

Tabelle 34: Confidencewerte in Studie 2 (M und SD)

Gruppe	Confidence Forecaster	Confidence Algorithmus	Differenz	Paardifferenzentest
CA	3,52 (0,99)	3,76 (0,82)	-0,24 (0,96)	t(266) = -4,160, p < 0,001
PA	3,46 (1,01)	3,64 (0,86)	-0,18(0,97)	t(284) = -3,111, p = 002
CBP	3,54 (0,99)	3,79 (0,81)	-0,26 (1,00)	t(269) = -4,196, p < 0,001
NBP	3,47 (1,13)	3,78 (0,84)	-0,31 (1,01)	t(282) = -5,115, p < 0,001
CGP	3,42 (0,98)	3,66 (0,79)	-0,24 (0,94)	t(272) = -4,158, p < 0,001
NGP	3,21 (0,97)	3,57(0,80)	-0,36 (0,99)	t(291) = -6,185, p < 0,001

### 5.3.2.4 Wirkung der Treatments und Selbstscheidung der Teilnehmer

Die Gruppen, die RPI erhalten haben, wurden nach ihrer Einschätzung zum Einsatz des Algorithmus bei ihren Peers gefragt. Die Mittelwerte der Antworten in Tabelle 35 liegen zwischen 61 % und 64 %. Das heißt, die Teilnehmer gehen davon aus, dass

<sup>685</sup> McCarty 1986, S. 844 zeigen diese Wirkung ebenfalls.

## Studie 2

sich über 60 % der anderen MTurker für den Algorithmus entschieden haben. Werden diese Einschätzungen mit Tabelle 36 und den Werten zur Algorithmen-Aversion innerhalb der Gruppen verglichen, zeigt sich, dass die Teilnehmer in den PA und der NGP den Einsatz von Algorithmen bei ihren Peers überschätzen. Für die NBP hingegen schätzen die Teilnehmer den Einsatz von Algorithmen fast genauso ein, wie er später auch in ihrer Gruppe vorliegt. Die realistische Einschätzung der Teilnehmer zur Algorithmen-Aversion ihrer Peers zeugt davon, dass sie sich mit dem Treatment auseinandergesetzt haben und wird daher als überzeugende Wirksamkeit des Treatments gewertet.

Tabelle 35: Einschätzung zum Algorithmen-Einsatz der Peers

<b>Gruppe</b>	<b>Mittelwert (SD)</b>
<b>PA (n = 285)</b>	62,61 (21,34)
<b>NBP (n = 283)</b>	63,98 (21,25)
<b>NGP (n = 292)</b>	61,01 (20,41)

Innerhalb der Gruppe mit negativer RPI variiert die Einschätzung zu den Peers in Abhängigkeit von der Leistung der Teilnehmer leicht. Sie liegt in der NGP bei 61,01 und bei der NBP bei 63,98. Diese Differenz ist nicht statistisch signifikant ( $t(573) = 1,847$ ,  $p = 0,065$ ), lässt sich dennoch als leichte Tendenz interpretieren, in der bei NGP den Peers tendenziell eine stärker ausgeprägte Algorithmen-Aversion zugesprochen wird als bei NBP.

### 5.3.2.5 Entscheidung im incentivierten Forecast

Die entscheidende Grundlage für die Hypothesentests liefert Tabelle 36. Im Falle von positiver RPI wächst die Wahl zum eigenen Forecast von 47,2 % auf 57,9 % an (CA zu PA). Es liegt demnach eine große Algorithmen-Aversion vor. Bei schlechter Leistung hingegen sind wiederum deutliche Unterschiede in der Präferenz für den eigenen Forecast zu sehen (CBP zu NBP). Während sich ohne negative RPI noch 46,7 % der Teilnehmer für den eigenen Forecast entscheiden, wählen bei negativer RPI bei schlechter Leistung lediglich 37,8 % der Teilnehmer ihren eigenen Forecast. Es liegt demnach eine geringe Algorithmen-Aversion vor. Für das Vorliegen von guter Leistung bei negativer RPI sind marginale Unterschiede zwischen den Gruppen festzustellen (CGP zu NGP).

Tabelle 36: Entscheidung im incentivierten Forecast in Studie 2

Gruppe	Wahl auf Forecaster %	Wahl auf Algorithmus in %
CA (n = 267)	47,2	52,8
PA (n = 285)	57,9	42,1
CBP (n = 270)	46,7	53,3
NBP (n = 283)	37,8	62,2
CGP (n = 273)	50,5	49,5
NGP (n = 292)	53,4	46,6

## 5.4 Hypothesentests und Beurteilung der empirischen Ergebnisse

### 5.4.1 Hypothesentests

In Tabelle 37 sind die Hypothesentests zu H5 und H6 dargestellt. Für die Hypothesentests wird ein einseitiger Chi-Quadrat-Test gerechnet. In H5 wird die Verteilung zwischen CA und PA und in H6 wird die Verteilung zwischen CBP und NBP getestet. Die jeweils untersuchten Gruppen referenzieren auf die Nomenklatur in Tabelle 26. Beide Hypothesen werden angenommen. Das bedeutet, dass sich bei positiver RPI die Algorithmen-Aversion signifikant erhöht (CA und PA). Bei negativer RPI und dem Vorliegen von schlechter Forecastinggenauigkeit, senkt sich die Algorithmen-Aversion signifikant (CBP und NBP).

Tabelle 37: Hypothesentests Studie 2

Hypothese	$\chi^2$	df	p	n
H5	6.336	1	0,006 <sup>686</sup>	552
H6	4,446	1	0,018 <sup>687</sup>	553

Der Hypothesentest zu H5 bzw. H6 basiert auf der Kreuztabelle in Tabelle 38 bzw. Tabelle 39. In Tabelle 38 zeigt sich die deutliche Verschiebung hin zum eigenen Forecast, während Tabelle 39 die Verschiebung hin zum Forecast des Algorithmus deutlich macht. Die Interpretationen zu H5 bzw. H6 finden sich in Abschnitt 0.

<sup>686</sup> Die Effektstärke von H5 mit n = 552, wird mit Cramer's-V = 0,107 angegeben. Es liegt demnach ein kleiner Effekt vor.

<sup>687</sup> Die Effektstärke von H6 mit n = 553, wird mit Cramer's-V = 0,090 angegeben. Es liegt demnach ein kleiner Effekt vor.

## Studie 2

Tabelle 38: Kreuztabelle H5

		Experimentalgruppe		
		CA	PA	Zeilensumme
Entscheidung	Wahl auf Forecaster	126 (47,2 %)	165 (57,9 %)	291 (52,7 %)
	Wahl auf Algorithmus	141 (52,8 %)	120 (42,1 %)	261 (47,3 %)
Spaltensumme		267 (100 %)	285 (100 %)	552 (100 %)

Tabelle 39: Kreuztabelle H6

		Experimentalgruppe		
		CBP	NBP	Zeilensumme
Entscheidung	Wahl auf Forecaster	126 (46,7 %)	107 (37,8 %)	233 (42,1 %)
	Wahl auf Algorithmus	144 (53,3 %)	176 (62,2 %)	320 (57,9 %)
Spaltensumme		270 (100 %)	283 (100 %)	553 (100 %)

### 5.4.2 Interpretationen der Ergebnisse und Zusatzanalysen

#### 5.4.2.1 Zusatzanalysen und Diskussion H5

In Hypothese H5 wurde prognostiziert, dass positive RPI zur Performance von Forecastern die Algorithmen-Aversion erhöht. Die Hypothese konnte bestätigt werden. Abbildung 51 macht deutlich, inwiefern sich die Experimentalgruppen in ihrer Algorithmen-Aversion unterscheiden. Mit lediglich 42,1 % Teilnehmer, die sich im incentivierten Forecast für den Algorithmus entschieden haben, liegt hier der niedrigste Wert über alle Experimentalgruppen dieser Untersuchung hinweg vor.<sup>688</sup> Die Bestärkung der Teilnehmer aus positiver RPI für ihre Leistung findet sich daher in der Algorithmen-Aversion wieder.

<sup>688</sup> Dies gilt für Studie 1 und Studie 2.

## Studie 2

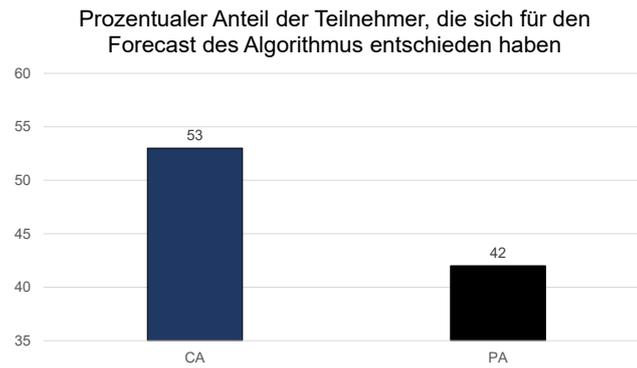


Abbildung 51: Algorithmen-Aversion bei positiver RPI<sup>689</sup>

Tabelle 40 weist die Confidencewerte und deren Unterscheidung zwischen der CA und der PA auf. Dabei zeigt sich, dass der Mittelwert in der CA sowohl für die Confidence in den eigenen Forecast als auch die Confidence in den Algorithmus durchweg über denen der PA liegt. Es findet demnach eine leichte Reduktion der Confidencewerte durch positive RPI statt. Zur Interpretation muss gesagt werden, dass sich dies beispielsweise mit der Untersuchung von McCarty (1986) deckt,<sup>690</sup> jedoch im Widerspruch zum aufhellenden Bild zur Wahrnehmung der eigenen Leistung durch positive RPI steht.<sup>691</sup> Das positive RPI verringert die Confidence in den Algorithmus stärker als die Confidence in den eigenen Forecast. Das heißt es bestätigt die allgemeine Tendenz der Wirkung von Leistungsfeedback in der Algorithmen-Aversion, bei der es zu einer Abwertung des Forecasts des Algorithmus kommt.

Tabelle 40: Confidencewerte zu H5 (M und SD)

	CA	PA	Differenz	Ungepaarter t-Test
<b>Confidence Algorithmus</b>	3,76 (0,82)	3,64 (0,86)	0,12	t(550) = 1,692, p = 0,091
<b>Confidence Forecaster</b>	3,52 (0,99)	3,46 (1,01)	0,06	t(550) = 0,671, p = 0,503

Da der t-Test der Confidence in den Algorithmus nahe an der Signifikanzgrenze liegt, werden die Confidencewerte in Abbildung 52 und Abbildung 53 zur besseren Interpretation grafisch gegenübergestellt. Der Vergleich von Abbildung 52 mit Abbildung 53 zeigt, dass durch die positive RPI die Confidence in den Algorithmus leicht

<sup>689</sup> Eigene Darstellung.

<sup>690</sup> Vgl. McCarty 1986, S. 840.

<sup>691</sup> Vgl. Kuhnen und Tymula 2012, S. 94.

## Studie 2

sinkt, jedoch sonst keinen bedeutenden Einfluss hat. Für die Confidence in den eigenen Forecast sinkt der Mittelwert ebenfalls ab. Die Verteilung verändert sich dabei von einer leicht rechtsschiefen Verteilung hin zu einer leicht linksschiefen Verteilung.

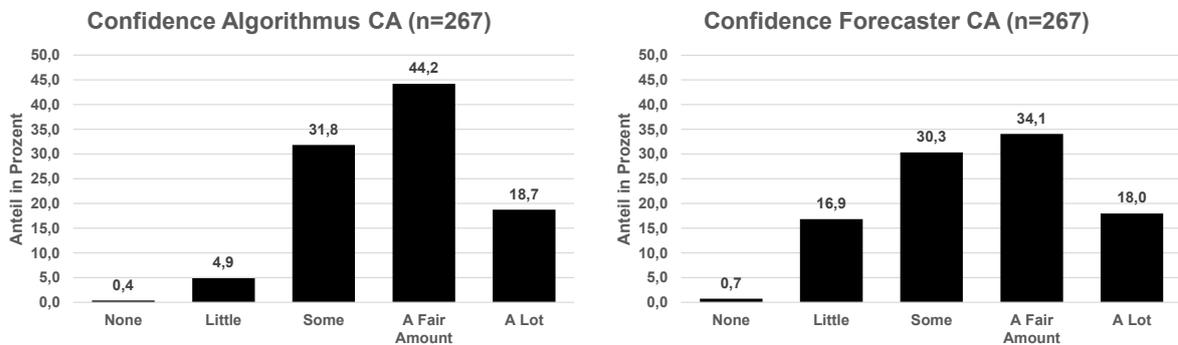


Abbildung 52: Verteilung der Confidencewerte in der CA<sup>692</sup>

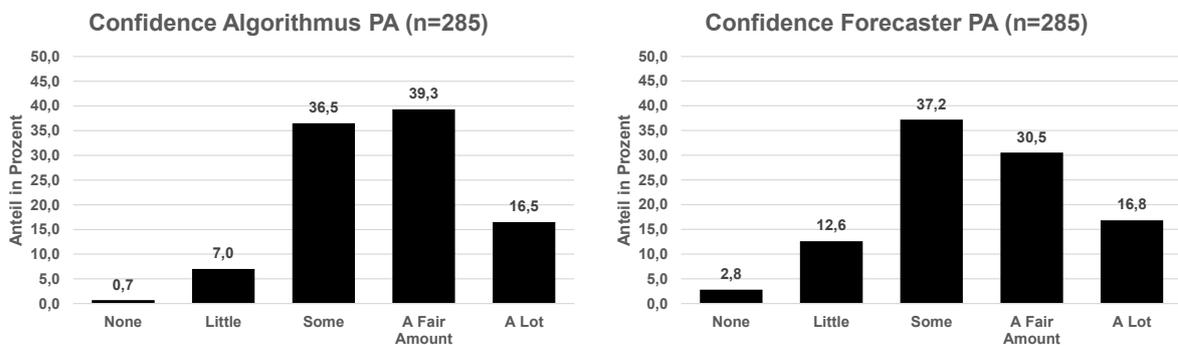


Abbildung 53: Verteilung der Confidencewerte in der PA<sup>693</sup>

Begründet werden kann die Wirkung von positiver RPI primär durch die Informationsfunktion von RPI. Es wird angenommen, dass die Wirkung des positiven Feedbacks die Teilnehmer demnach auf der untersten der hierarchischen Wirkebenen der FIT erreicht. Im Wesentlichen wurde ihre Arbeitsweise, also ihr Lernprozess, auf der Lernebene verstärkt. Entsprechend der aus der Theorie abgeleiteten Erwartung, hat sich der Aufmerksamkeitsfokus nicht von der Lernebene entfernt. Die Teilnehmer hinterfragen ihre Leistung durch das positive RPI stärker, woraus eine leichte Reduktion der Confidence resultiert. Zeitgleich finden sie in Summe jedoch eine Bestätigung ihrer Leistung bzw. Leistungserbringung und verfolgen damit vermehrt den eigenen Forecast und nicht den des Algorithmus. Ein Indiz hierfür stellt die leicht, jedoch nicht signifikante Verbesserung der Leistung der Teilnehmer in der PA gegenüber der CA dar

<sup>692</sup> Eigene Darstellung.

<sup>693</sup> Eigene Darstellung.

## Studie 2

(siehe Tabelle 41). In Summe sinkt die Confidence der Teilnehmer also durch positives RPI leicht ab, während ihre Leistung leicht ansteigt. Diese Effekte sind jedoch nicht signifikant.

Tabelle 41: Performancevergleich CA und PA (M und SD)

	CA	PA	Differenz	Ungepaarter t-Test
<b>Performance Incentiviert</b>	9,39 (8,72)	9,00 (8,92)	0,39	t(550) = 0,518, p = 0,604

Zusammenfassend wird festgehalten, dass Feedbackempfänger in der Regel eine Präferenz für positives Feedback aufweisen und es bereitwillig akzeptiert und als zutreffend wahrgenommen wird.<sup>694</sup> Die Algorithmen-Aversion wird dadurch signifikant erhöht. Positives Peer-Feedback führt zu einer verstärkten Überzeugung der eigenen Leistung gegenüber der Leistung des Algorithmus.<sup>695</sup> Dies wiederum führt zur Überzeugung, auch in Zukunft – also im Falle der vorliegenden Studie: dem incentivierten Forecast – eine gute Leistung zu erbringen. Der Einfluss von positiver RPI auf die steigende Algorithmen-Aversion konnte deutlich gezeigt werden. Mit der Fokussierung auf die Lernebene entsteht womöglich ein Fokus auf die eigene Leistung und damit eine verstärkte Abkehr von der Wahl des Algorithmus. Es ist daher anzunehmen, dass positive RPI einen, entsprechend der FIT prognostizierten, Motivationseffekt hat. Auch der aus der SCT vorhergesagte Anstieg der Algorithmen-Aversion tritt ein. Es ist jedoch keine Verbesserung des Selbstbilds in Form von mehr Confidence in die eigene Leistung entstanden und es kann interpretiert werden, dass allein die Informationsfunktion aus positiver RPI negativ auf die Interaktion mit Algorithmen wirkt.

### 5.4.2.2 Zusatzanalysen und Diskussion H6

Für Hypothese H6 wurde prognostiziert, dass negative RPI zur Forecastinggenauigkeit von leistungsschwachen Forecastern die Algorithmen-Aversion senkt. Die Hypothese konnte bestätigt werden. Abbildung 54 (links) zeigt, inwiefern sich die Experimentalgruppen unterscheiden. Bei 62,2 % der Teilnehmer, die sich bei negativer RPI und schlechter eigener Leistung im incentivierten Forecast für den Algorithmus entschied-

<sup>694</sup> Vgl. Holderness et al. 2017, S. 35.

<sup>695</sup> Vgl. Kuhnen und Tymula 2012, S. 94.

## Studie 2

den haben, liegt hier unter Leistungsfeedback der höchste Wert über alle Experimentalgruppen hinweg vor.<sup>696</sup> Abbildung 54 (rechts) zeigt analog, wie gering der Einfluss der negativen RPI bei guter Leistung ist.

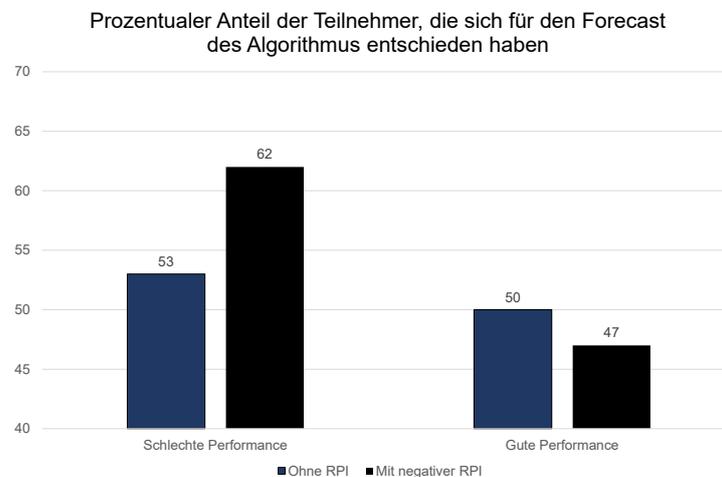


Abbildung 54: Algorithmen-Aversion bei negativer RPI<sup>697</sup>

Die aus der Theorie heraus notwendige Trennung in gute und schlechte Performer zeigt ein differenziertes Bild der Algorithmen-Aversion.<sup>698</sup> Für die Gruppen mit guter Leistung ergibt sich die Wahl zum Algorithmus wie folgt: Bei NGP liegt augenscheinlich eine ähnliche Verteilung der Wahl zum Algorithmus vor wie in der CGP. Bei der NBP gegenüber der CBP verschiebt sich die Häufigkeit hin zur Wahl des Algorithmus signifikant. Dies bestätigt die Literatur zu RPI und lässt sich konform damit argumentieren.<sup>699</sup> Die negative RPI lässt bei schlechter Leistung einen überzeugenden Rückschluss auf die eigene Leistung und damit auf die eigenen Fähigkeiten zu. Dieser Rückschluss erzeugt eine ungewollte Transparenz zur Leistung beim Feedbackempfänger.<sup>700</sup> Durch die Wahl des Algorithmus, kann dieser *unangenehme* Rückschluss reduziert werden und eine alternative Arbeitsweise zur besseren Aufgabenerfüllung verfolgt werden. Flankiert wird dieses Verhalten durch die Möglichkeit, mit dem Algorithmus das Incentive zu maximieren.

Tabelle 42 zeigt die Confidencewerte und wie sich diese zwischen den CBP und NBP verhalten. Dabei ist ersichtlich, dass es kaum einen Unterschied zwischen den

<sup>696</sup> Dies gilt für Studie 1 und Studie 2.

<sup>697</sup> Eigene Darstellung.

<sup>698</sup> Wenn keine Unterscheidung in gute und schlechte Performance der Teilnehmer gemacht wird, dann lässt sich auch kein Einfluss durch negative RPI auf die Algorithmen-Aversion feststellen ( $\chi^2(1) = 0,929$ ,  $p = 0,335$ ,  $n = 1118$ ).

<sup>699</sup> Vgl. Arnold et al. 2018, S. 1; Ashraf et al. 2014, S. 44.

<sup>700</sup> Vgl. Arnold et al. 2018, S. 1.

## Studie 2

Gruppen gibt. Der Einfluss auf die Algorithmen-Aversion muss daher bei der NBP durch die aus der SCT bekannten Informationsfunktion der RPI erklärt werden. Die von der FIT und SCT zusätzlich möglichen Erklärungsansätze über eine Wirkung auf das Selbst und die Confidence in die eigene Leistung werden hier nicht bestätigt. Jedoch erscheint eine bessere Einordnung der schlechten Leistung, durch eine zusätzliche Information zu dieser, ein starker Faktor zur Senkung der Algorithmen-Aversion zu sein.

Die FIT legt nahe, dass durch die negative RPI und damit das doppelt negative Feedback zur Leistung (Leistungsfeedback aus Übungsforecasts und negative RPI) eine tendenzielle Abkehr von der Lernebene beim Empfänger stattfindet. Dies hat zur Folge, dass die Intensivierung des Arbeitseinsatzes (*working harder*) nicht weiterverfolgt und auf der Lern- und Motivationsebene versucht wird, eine neue Strategie zu finden, um die Aufgabe zu erfüllen (*working smarter*).<sup>701</sup> Es ist aus Sicht der FIT zu argumentieren, dass zwar kein stärkeres Hinterfragen der eigenen Leistung, jedoch eine Verschiebung auf die Motivationsebene zur Problemlösung eintritt. Es kommt demnach zu einer Verschiebung auf die Motivationsebene wobei der Aufgabenebenen-Fokus erhalten bleibt und ein Wechsel der Strategie von *working harder* zu *working smarter* wird vollzogen.

Tabelle 42: Vergleich der Confidencewerte bei CBP und NBP (M und SD)

	CBP	NBP	Differenz	Ungepaarter t-Test
<b>Confidence Algorithmus</b>	3,79 (0,81)	3,78 (0,84)	0,02	t(551) = 0,216, p = 0,829
<b>Confidence Forecaster</b>	3,54 (0,99)	3,47 (1,13)	0,07	t(551) = 0,742, p = 0,458

Der Vergleich der Forecastinggenauigkeit zeigt ebenfalls keine signifikanten Änderungen (Tabelle 43). Bei der Interpretation von Tabelle 43 sei darauf verwiesen, dass die dargestellten Gruppen jeweils durch einen Median-Split entstanden sind und unter RPI generell eine leichte Verbesserung der Leistung zu erwarten war.

<sup>701</sup> Vgl. Kluger und DeNisi 1996, S. 263.

## Studie 2

Tabelle 43: Vergleich der Forecastinggenauigkeit CBP und NBP (M und SD)

	CBP	NBP	Differenz	Ungepaarter t-Test
<b>Performance Incentiviert</b>	12,76 (9,79)	11,72 (9,73)	1,04	t(551) = 1,251, p = 0,212

Es wird festgehalten, dass die Algorithmen-Aversion durch die Einführung von RPI bei schlechter Performance bedeutend gesenkt wurde. Da in der Gegenüberstellung der CBP mit der NBP ansonsten sehr ähnliche Werte hinsichtlich der Confidence und der Forecastinggenauigkeit vorliegen, wird weiter untersucht, worin sich das Verhalten der NBP begründen lässt. Dies wird durch die Gegenüberstellung der NGP mit der NBP und der CGP mit der CBP erreicht, also jeweils der guten mit der schlechten Performance.

### 5.4.2.3 Zusatzanalysen und Diskussion zu guter vs. schlechter Performance

Im Folgenden wird gezeigt, welche Auswirkungen durch die Aufteilung in gute und schlechte Performer entstehen. Tabelle 44 zeigt die Confidencewerte unter negativer RPI. Tabelle 45 zeigt die Confidencewerte ohne RPI (Kontrollgruppen) differenziert nach guter und schlechter Performance.

Tabelle 44: Vergleich der Confidencewerte NGP und NBP (M und SD)

	NGP	NBP	Differenz	Ungepaarter t-Test bzw. Welch-Test
<b>Confidence Algorithmus</b>	3,57 (0,80)	3,78 (0,84)	-0,21	t(573) = -3,098, p = 0,002
<b>Confidence Forecaster</b>	3,21 (0,97)	3,47 (1,13)	-0,22	t(553) <sup>702</sup> = -3,011, p = 0,003

Tabelle 45: Vergleich der Confidencewerte CGP und CBP (M und SD)

	CGP	CBP	Differenz	Ungepaarter t-Test
<b>Confidence Algorithmus</b>	3,66 (0,78)	3,79 (0,81)	-0,13	t(541) = -1,937, p = 0,053
<b>Confidence Forecaster</b>	3,42 (0,98)	3,47 (1,13)	-0,12	t(541) = -1,374, p = 0,170

<sup>702</sup> Varianzen sind nicht gleich, weshalb ein Welch-Test verwendet wurde. Siehe dazu auch Rasch et al. 2011.

## Studie 2

Die Teilnehmer haben sowohl mit als auch ohne negative RPI bei schlechter Leistung (CBP und NBP) sowohl mehr Confidence in ihren Forecast als auch in den Forecast des Algorithmus. Diese höheren Confidencewerte sind jedoch nur beim Vorliegen von negativer RPI signifikant gegenüber den Teilnehmern mit guter Performance (siehe Tabelle 44 und Tabelle 45). Eine signifikante Veränderung beider Confidencewerte erfordert eine vertiefte Analyse hinsichtlich der daraus resultierenden Algorithmen-Aversion. Dazu werden die Ergebnisse, im Kontext der Algorithmen-Aversion, in die Entscheidung im incentivierten Forecast gesetzt: Ein Chi-Quadrat-Tests zwischen NGP und NBP sowie CGP und CBP unterstützt die Interpretation (siehe Tabelle 46). Die Algorithmen-Aversion zwischen guten und schlechten Performern unter negativer RPI ist hochsignifikant ( $\chi^2(1) = 14,120$ ,  $p < 0,001$ ,  $n = 575$ ). Wird die Algorithmen-Aversion hingegen ohne negative RPI geprüft (CGP und CBP), liegt kein Unterschied zwischen den Gruppen vor ( $\chi^2(1) = 0,819$ ,  $p = 0,365$ ,  $n = 543$ ). Tabelle 46 zeigt diese Zusammenhänge. Das bedeutet, die Unterschiede im Verhalten der Teilnehmer zwischen Kontrollgruppe und negativer RPI ist ausgelöst durch die negative RPI – und wirkt besonders bei schlechter Performance.

Tabelle 46: Chi-Quadrat-Test zur Algorithmen-Aversion bei guter und schlechter Performance

Gruppen	$\chi^2$	df	p	n
NGP und NBP	14.120	1	< 0,001	575
CGP und CBP	0,819	1	0,265	543

Die starke Algorithmen-Aversion zwischen der NGP und NBP kann konsistent mit Studie 1, *bei signifikant unterschiedlichen Confidencewerten*, durch die Mediationsanalyse vertieft werden.<sup>703</sup> Da sowohl die Confidence in den Algorithmus als auch die Confidence in den eigenen Forecast signifikant steigt, ist a priori eine Diskussion zur Richtung des erwarteten Effekts der Mediation notwendig.<sup>704</sup>

Die bisherigen Ausführungen legen die Annahme zugrunde, dass die Zunahme (Abnahme) der Confidence in den eigenen Forecast zu einer Zunahme (Abnahme) der Algorithmen-Aversion führt. Hier sei auf Studie 1 und den Einfluss des Zeitdrucks verwiesen. Zeitdruck senkt die Confidence in den eigenen Forecast und damit die Algorithmen-Aversion. Es wurde ebenfalls vielfach gezeigt, dass eine sinkende (steigende)

<sup>703</sup> Siehe Abschnitt 4.4.3.1.

<sup>704</sup> Vgl. Preacher und Hayes 2004, S. 717.

## Studie 2

Confidence in den Algorithmus die Algorithmen-Aversion erhöht (senkt), siehe dazu die Ausführungen in Abschnitt 4.4.3 in Studie 1 und Dietvorst et al. (2015). Diese empirisch begründeten Zusammenhänge sind nicht konform mit dem Chi-Quadrat-Test zur Algorithmen-Aversion unter negativer RPI in Tabelle 46 und den dazugehörigen Confidencewerten in Tabelle 44. Sowohl die Confidence in den eigenen Forecast als auch die Confidence in den Algorithmus steigt signifikant von guter zu schlechter Performance (Tabelle 44). Die Algorithmen-Aversion sinkt signifikant von guter zu schlechter Performance, denn bei schlechter eigener Leistung und negativer RPI wählen die Teilnehmer bereitwillig den Algorithmus (siehe auch Abbildung 54). Aus den Chi-Quadrat-Tests zur Algorithmen-Aversion hätte in Tabelle 44 folgen müssen, dass die Confidence in den Algorithmus steigt und/oder die Confidence in den eigenen Forecast sinkt. Die Confidence in den eigenen Forecast steigt jedoch.

Dieser vermeintliche Widerspruch wird durch die Mediationsanalyse vertieft. Es wird dazu eine binäre Mediationsanalyse mit 95 % Konfidenzintervallen um den indirekten Effekt gerechnet. Die Stichprobengröße beträgt 575 Teilnehmer. Dies umfasst die NGP und die NBP. Die abhängige Variable ist die Entscheidung der Teilnehmer im incentivierten Forecast für oder gegen den Algorithmus (Y).<sup>705</sup> Die unabhängige Variable ist, ob eine gute oder schlechte Performance vorliegt (X).<sup>706</sup> Der Mediator ist die Confidence in den Algorithmus (Abbildung 55) bzw. die Confidence in den eigenen Forecast (Abbildung 56) (M). Beide Mediatoren sind dabei signifikant:

Mit dem indirekten Effekt  $ab = 0,1294$ , 95 % CI [0,0415, 0,2392],  $n = 575$ , mediiert die Confidence in den Algorithmus die Algorithmen-Aversion, wenn Forecaster schlechte Leistungen erbringen und ihnen negative RPI gegeben wird (siehe Abbildung 55). Für den a-Pfad gilt ein Koeffizient von 0,2123 mit  $p = 0,002$ , für den b-Pfad gilt 0,6095 mit  $p < 0,001$ . Für den direkten Effekt gilt, wenn auf die Confidence in den Algorithmus hin kontrolliert wird, ist  $c' = 0,7243$ ,  $p < 0,001$ .

---

<sup>705</sup> Kodiert mit: 0 = Wahl auf eigenen Forecast, 1 = Wahl auf Algorithmus.

<sup>706</sup> Kodiert mit: 0 = gute Leistung, 1 = schlechte Leistung.

## Studie 2

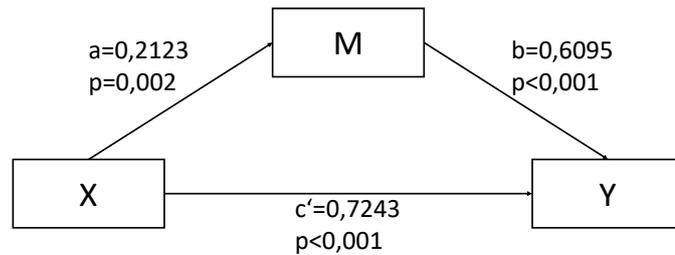


Abbildung 55: Mediationsanalyse RPI (1/2)<sup>707</sup>

Mit dem indirekten Effekt  $ab = -0,0741$ , 95 % CI  $[-0,1510, -0,0182]$ ,  $n = 575$ , mediiert die Confidence in den eigenen Forecast die Veränderung der Algorithmen-Aversion, wenn Forecaster schlechte Leistungen erbringen und negative RPI gegeben wird (Abbildung 56). Für den a-Pfad gilt  $0,2645$  mit  $p = 0,003$ , für den b-Pfad gilt  $-0,2803$  mit  $p < 0,001$ . Für den direkten Effekt gilt, wenn auf die Confidence in den eigenen Forecast kontrolliert wird, ist  $c' = 0,5426$ ,  $p = 0,002$ .

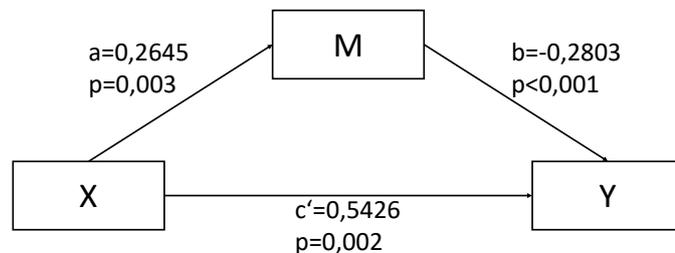


Abbildung 56: Mediationsanalyse RPI (2/2)<sup>708</sup>

Der Wechsel im Vorzeichen des Mediators in Abbildung 56 ist in der Wirkrichtung begründet. Die signifikante Mediation ist auch dann gültig, wenn direkte Effekte nicht signifikant sind oder, wie in diesem Fall, mit einem anderen Vorzeichen versehen sind als der indirekte Effekt.<sup>709</sup> Hayes (2018) begründet dies durch die Komplexität realer Beziehungen und plädiert für sowohl eine integrierte Betrachtung von direktem und indirektem Effekt als auch eine isolierte Betrachtung des indirekten Effekts.<sup>710</sup> Eine Mediation, bei der der direkte und der indirekte Effekt nicht das gleiche Vorzeichen tragen, wird als inkonsistente Mediation bzw. *opposing mediated effect* bezeichnet.<sup>711</sup> MacKinnon et al. (2007) halten dabei fest: „Inconsistent mediator effects may be es-

<sup>707</sup> Eigene Darstellung.

<sup>708</sup> Eigene Darstellung.

<sup>709</sup> Vgl. Hayes 2018, S. 80.

<sup>710</sup> Vgl. Hayes 2018, S. 81.

<sup>711</sup> Vgl. MacKinnon et al. 2007, S. 602.

pecially critical in evaluating counterproductive effects of experiments, where the manipulation may have led to opposing mediated effects“<sup>712</sup>. Für die Confidence in den eigenen Forecast liegt ein *opposing mediated effect* vor (Abbildung 56). Wenn auf die Confidence in den eigenen Forecast hin kontrolliert wird, wird der direkte Effekt ( $c'$ ) kleiner. Das heißt, die theoretischen Annahmen zu den Zusammenhängen in der Algorithmen-Aversion sind weiterhin gültig. Die Algorithmen-Aversion sinkt von guter zu schlechter Leistung durch eine Zunahme der Confidence in den Algorithmus – dieser Effekt wird *abgemildert*, jedoch nicht eliminiert, durch eine Zunahme der Confidence in den eigenen Forecast.

Rückblickend auf die Hypothese H6 kann damit für die empirische Untersuchung wie folgt argumentiert werden. Aus der Literatur geht hervor, dass für Forecaster mit schlechter Performance eine höhere Sensibilität für RPI erwartet werden kann.<sup>713</sup> Dieser Effekt hätte demnach auf die Confidence in den eigenen Forecast wirken sollen. Der Effekt auf die Confidence lässt sich jedoch zwischen CBP und NBP nicht nachweisen (Tabelle 42). Die Erklärung der Zusammenhänge von Algorithmen-Aversion und Confidence kann in dieser Untersuchung zwar durch die Mediationsanalyse zwischen NGP und NBP unterstützt werden, jedoch kann keine abschließende Erklärung erfolgen. Eine Möglichkeit zur Interpretation ist es, dass Forecaster grundsätzlich versuchen, die Wirkung von negativer RPI nicht auf ihre eigene Leistung zu projizieren (*Self-Serving Bias*).<sup>714</sup> Das heißt, wenn die RPI als glaubhaft wahrgenommen wird (bei schlechter Performance), ist die einzige Ausweichstrategie die, dass die Teilnehmer der NBP die Annahme treffen, dass ihre Peers den Algorithmus wählen, weil dieser eine gute Performance hat. Die Bewertung der eigenen Leistung rückt damit in den Hintergrund und kann zum Zwecke des Self-Serving Bias ebenfalls auf den Algorithmus übertragen werden. Die Algorithmen-Aversion sinkt damit entsprechend, ohne dass die Confidence in den eigenen Forecast sinkt.

Für gute Performer (bei negativem RPI) ist dieser Effekt weniger stark, da die Forecaster das Feedback als weniger glaubhaft einstufen und damit weniger über die Arbeit ihrer Peers reflektieren. Es findet demnach keine Aufwertung des Algorithmus

---

<sup>712</sup> MacKinnon et al. 2007, S. 602.

<sup>713</sup> Vgl. Falk und Ichino 2006, S. 39.

<sup>714</sup> Vgl. Bénabou und Tirole 2002, S. 871.

durch das antizipierte Verhalten der Peers statt. Die insgesamt niedrigeren Confidencewerte können einzig aus der widersprüchlichen RPI bei der NGP zur tatsächlichen Leistung stammen, wodurch eine generelle Abwertung der Aufgabe bzw. des Experiments entstanden sein könnte.

Die These, dass ein Self-Serving Bias bei NBP vorliegt, kann mit der *Einschätzung zu den Peers* unterstützt werden. Tabelle 35 zeigt zur Einschätzung vom Einsatz des Algorithmus bei den Peers keine signifikanten Unterschiede zwischen der NBP und der NGP. Jedoch korreliert für NBP die Peereinschätzung mit  $r = 0,326$ ,  $p < 0,001$  mit der Confidence in den Algorithmus. Das heißt, je mehr die Teilnehmer an den Einsatz des Algorithmus bei den Peers glauben, desto höher ist die Confidence in den Algorithmus. Für die NBP gilt dieser Zusammenhang ebenfalls, jedoch mit  $r = 0,175$ ,  $p < 0,001$  deutlich schwächer.

### 5.4.3 Schlussfolgerungen

Studie 2 erforscht, wie sich die Algorithmen-Aversion verhält, wenn Forecaster RPI erhalten. RPI ergänzt das Leistungsfeedback, das die Algorithmen-Aversion auslöst, um eine weitere Feedbackdimension. Es wurde sowohl für positive wie auch negative RPI gezeigt, dass eine signifikante Wirkung auf die Algorithmen-Aversion besteht (siehe Tabelle 47). Da RPI die Algorithmen-Aversion sowohl senken, wie auch erhöhen kann, kann sie als *zweischneidiges Schwert* interpretiert werden.

Tabelle 47: Zusammenfassung der Hypothesentests in Studie 2

Hypothese	Prognostizierte Wirkung	Hypothesentest
H5: Positive RPI zur Forecastinggenauigkeit erhöht die Algorithmen-Aversion.	↑	<input checked="" type="checkbox"/>
H6: Negative RPI zur Forecastinggenauigkeit reduziert bei leistungsschwächeren Forecastern die Algorithmen-Aversion.	↓	<input checked="" type="checkbox"/>

Kluger und DeNisi (1996) und DeNisi und Kluger (2000) empfehlen, dass effektives Feedback nur in Kombination mit einer Zielsetzung verwendet werden sollte. Feedback, das sich auf zuvor festgelegte, aufgabenbezogene Ziele bezieht, lenkt die Aufmerksamkeit eher auf die Aufgabe als auf das Selbstbild. Dies bestätigt die vorliegende Untersuchung teilweise. Die Forecaster in Studie 2 verhalten sich dem Feedback durch RPI konform und passen ihre Arbeitsweise diesem an. Das heißt, dass bei

## Studie 2

positiver RPI die Algorithmen-Aversion steigt und bei negativer RPI (und schlechter Performance) die Algorithmen-Aversion sinkt.

Dies ist für die Praxis von Relevanz, denn RPI kann damit Forecaster die keine gute Leistung aufweisen unterstützen. Diese Forecaster sind bei nur einer Feedbackdimension (Leistungsfeedback, siehe Abschnitt 2.1.3) unzureichend in der Lage, ihre Leistung richtig einzuschätzen. Das negative RPI zeigt ihnen auf, dass sie gegenüber ihren Peers ebenfalls keine gute Leistung erbringen. Ähnlich der Betrachtung zur Arbeit mit Algorithmen unter Zeitdruck, unterstützt RPI den Ausweg aus der unzureichenden eigenen Leistung und lenkt die Forecaster hin zur Verwendung des Algorithmus. Demgegenüber sind Forecaster bei positiver RPI geneigt, ihre bestehende Arbeitsweise beizubehalten, und entscheiden sich tendenziell vermehrt gegen den Algorithmus.

Es lässt sich daher festhalten, dass RPI dann eingesetzt werden sollte, wenn ein Algorithmus eine etwas bessere Leistung als der Mittelwert der Forecastinggenauigkeit der Forecaster, die mit ihm arbeiten, erbringt. Für gute Forecaster, die wiederum etwas schlechter oder ähnlich gut wie der Algorithmus sind, entsteht aus der steigenden Algorithmen-Aversion kaum ein Nachteil bei der Vergütung, da sie auch mit ihrem Forecast eine hohe Präzision erreichen können. Forecastern, mit schlechter Leistung, wird hingegen durch RPI eine Hilfestellung zur Evaluation ihrer Leistung gegeben. Vor dem Hintergrund, dass RPI nicht einseitig gegeben werden kann und Forecaster in der Praxis zumeist in Teams arbeiten, werfen Falk und Ichino (2006) die Frage auf: „Should employees work in groups or alone? If they work in groups, how should low-productivity and high-productivity workers be optimally grouped?“<sup>715</sup> Diese Frage lässt sich basierend auf den empirischen Ergebnissen dieser Arbeit näherungsweise beantworten. Unter der Prämisse, dass Eingriffe der Forecaster in das Ergebnis des Algorithmus den Forecast in der Regel schlechter machen, sollten diese minimiert werden. Jedoch sollten sie zugelassen werden, wenn es starke und kompetente Eingriffe sind.<sup>716</sup> Das bedeutet, dass die Zusammenarbeit in Teams, mit Algorithmen und RPI so ausgestaltet sein sollte, dass einige wenige Top-Performer mit vielen Low-Performern zusammenarbeiten. Die Low-Performer profitieren von der Unterstützung des

---

<sup>715</sup> Falk und Ichino 2006, S. 55.

<sup>716</sup> Vgl. Fildes et al. 2009, S. 3.

## Studie 2

Algorithmus und nehmen diesen durch RPI bereitwillig an. Die Top-Performer wiederum haben die Fähigkeit Anpassungen am Algorithmus vorzunehmen, um diesen zu verbessern, wenn er nicht optimal arbeitet. Unter diesen Annahmen kann die Einführung von RPI in Teams nutzenstiftend erfolgen.

Burton et al. (2019) trafen theoretische Überlegungen zur Wirkung von RPI, die durch die Ergebnisse dieser Untersuchung teilweise verworfen werden müssen.<sup>717</sup> Zunächst haben sie in einer Art spieltheoretischer Überlegung angenommen, dass es bei der Einführung von RPI zu einer Abkehr vom Algorithmus, also einer starken Algorithmen-Aversion, kommt. Sie nahmen an, dass durch die antizipierte kollektive Wahl des Algorithmus sich der Einzelne nur besserstellen kann, wenn er nicht den Algorithmus wählt. Diesen Überlegungen kann mit zwei Argumente widersprochen werden. Zunächst antizipieren die Teilnehmer in Studie 2, dass etwas mehr als 60 % ihrer Peers den Algorithmus wählen. Den Teilnehmern ist demnach bewusst, dass eine gewisse Algorithmen-Aversion bei ihren Peers vorliegt (siehe Tabelle 35). Weiter haben Burton et al. (2019) nicht berücksichtigt, dass die Wirkung von RPI auf die Algorithmen-Aversion eine Berücksichtigung von positiver wie auch negativer RPI bedarf. Für negative RPI und schlechte Leistung treffen ihre Überlegungen nicht zu, da vor dem Hintergrund der individuellen Incentivierung, die Teilnehmer mit dem Algorithmus die Forecastinggenauigkeit verbessern und das Incentive maximieren können. Für positive RPI kann ihre Überlegung als zutreffend wahrgenommen werden und Feedbackempfänger fühlen sich durch die positive RPI einerseits bestätigt und andererseits womöglich herausgefordert, noch besser zu werden. Dies ist nur durch die Verbesserung der eigenen Leistung möglich.

Für die FIT ist zu bewerten, dass bei negativer RPI und schlechter Leistung eine tendenzielle Abkehr von der Lernebene stattfindet. Es wurde dazu argumentiert, dass zwar kein stärkeres Hinterfragen der eigenen Leistung, jedoch eine Verschiebung auf die Motivationsebene zur Problemlösung eintritt. Es kommt demnach zu einer Verschiebung auf die Motivationsebene, wobei der Aufgabenebenen-Fokus erhalten bleibt und ein Wechsel der Strategie von *working harder* zu *working smarter* vollzogen

---

<sup>717</sup> Vgl. Burton et al. 2019, S. 5.

## Studie 2

wird. Hierzu ergänzend wurde argumentiert, dass es durch positive RPI einen, entsprechend der FIT prognostizierten, Motivationseffekt gibt, der die eigene Leistung verstärkt und damit die Wahl des eigenen Forecasts fördert.

Eine abschließende Bewertung zur SCT kann für negative RPI und schlechte Performance durch den identifizierten *Self-Serving Bias* erfolgen. Der soziale Vergleich wird von schlechten Performern vermieden oder es werden alternative Erklärungen und Auswege gesucht.<sup>718</sup> Die Algorithmen-Aversion sinkt bei negativer RPI, da es durch den Algorithmus möglich ist, die als tendenziell persönlich wahrgenommene Kritik an der Leistung durch RPI mit der Wahl auf den Algorithmus abzuwenden. Eine bedeutsame Nebenbedingung zur Wirkung von RPI wird durch die *Theory of Downward Comparison* als Teilbereich der SCT formuliert.<sup>719</sup> Forecaster streben grundsätzlich nach einem positiven Selbstbild. Dieses Selbstbild kann sich jedoch durch negative Erfahrungen, wie unzureichende Leistungen im Beruf, verschlechtern. Durch RPI zu Peers, die schlechter abschneiden, kann wiederum ein aufhellendes, positives Selbstbild zur eigenen Leistung entstehen, wenn eine Überlegenheit zu diesen Peers wahrgenommen wird.<sup>720</sup> Es ist demnach irrelevant, wie gut ein Forecaster seine individuelle Leistung erbringt, da sich seine Art und Weise, die Leistung zu erbringen, verstärken wird, wenn durch positive RPI ein Abwärtsvergleich möglich ist.

### 5.4.4 Limitationen

Studie 2 unterlag in einigen Aspekten Limitationen. Zunächst konnte nicht vollständig sichergestellt werden, dass die Teilnehmer durch die Einteilung der Peer-Group nach Effort unterschiedliche Strategien verfolgt haben. Es ist denkbar, dass die Teilnehmer einen nicht linearen Zusammenhang von Zeitaufwand (*Effort*) und Leistung bzw. Forecastinggenauigkeit sehen. Konkret resultiert daraus, dass sich Teilnehmer durch schnelle Forecasts eine höhere Chance ausrechnen, in der Peer-Group besser abzuschneiden, da die Teilnehmer glauben in schnelleren Gruppen wurde ein geringerer Effort erbracht. Diese Annahme könnten Teilnehmer treffen, die glauben, dass Peers, die nur sehr wenig Zeit benötigen, sich durch die Aufgaben *durchklicken* und nicht gründlich nachdenken. Während hingegen in Gruppen mit großem Zeitaufwand mit

---

<sup>718</sup> Vgl. Bénabou und Tirole 2002, S. 871; Pyszczynski et al. 1985, S. 195.

<sup>719</sup> Vgl. Gibbons et al. 2002, S. 865; Wills 1981, S. 245.

<sup>720</sup> Vgl. Alicke und Govorun 2005, S. 85.

## Studie 2

Forecastern konkurriert wird, die sehr sorgfältig nachgedacht haben. Höherer Zeitaufwand in Verbindung mit besserer Leistung wurde bereits mehrfach nachgewiesen.<sup>721</sup> Vice versa ist es denkbar, dass die Teilnehmer annehmen, unter Peers mit hohem Zeitaufwand befänden sich weniger intelligente oder arbeitsscheue Forecaster. Das würde bedeuten, sie begeben sich freiwillig in diese Peer-Group. Ob es solche strategischen Überlegungen der Teilnehmer gab und ob sie einen Einfluss auf die Ergebnisse in Studie 2 haben, ist nicht endgültig auszuschließen.

Durch die Wirkung von RPI entsteht, selbst bei davon unabhängiger, individueller Leistungsvergütung, ein Wettbewerbsgedanke.<sup>722</sup> Dieser kann zu einem unerwünschten Verhalten führen, indem Forecaster beispielsweise ihr Wissen nicht mehr bereitwillig teilen.<sup>723</sup> Ein solches ungewolltes Verhalten kann aus unterschiedlichen Gründen entstehen. Die Basis dafür ist jedoch, dass ein Forecaster davon ausgeht, dass er sich durch dieses Wissen bzw. durch diese Fähigkeiten gegenüber seinen Peers besserstellen kann. Er wird das Wissen nicht teilen und er wird versuchen, seinen Wissensvorsprung intensiv zu erhalten bzw. auszubauen. Das in dieser Arbeit zugrunde liegende Experiment konnte solche Verhaltensweisen nicht berücksichtigen. Ergo sollte die künftige Forschung zur RPI und der Algorithmen-Aversion Forecaster nicht nur getrennt in Online-Experimenten untersuchen, sondern diesen eine Möglichkeit zur direkten Interaktion bzw. Beobachtung des gegenseitigen Verhaltens geben.

In der Studie 2 konnten keine wesentlichen Auswirkungen auf die Confidence in den eigenen Forecast durch RPI festgestellt werden. Dies legt nahe, dass die Teilnehmer der Studie keine wesentlichen Veränderungen auf ihr Selbstbild wahrgenommen haben. Die Wirkung der RPI resultierte demnach aus der Informationsfunktion von RPI. Dies legt wiederum nahe, dass die Lern- und Motivationsebene betroffen war, jedoch nicht die des Selbstbildes. Künftige Studien können die Wirkung von RPI in der Algorithmen-Aversion weiter vertiefen. Dabei können weitere Variablen zur Erklärung der Wirkung auf das Selbstbild und anderen Verhaltensweisen untersucht werden. Hier ist beispielhaft die Veränderung der Stichprobe hin zu Experimenten, in denen sich die Probanden persönlich kennen zu nennen. Hierdurch können die Effekte von

---

<sup>721</sup> Efendić et al. 2020, S. 104 schreiben hierzu: „Importantly, the more effort people believe others invest in completing relatively difficult tasks, whether in the form of time, physical exertion, pain, or money, the more positive the outcome of that effort is evaluated“.

<sup>722</sup> Vgl. Tafkov 2013, S. 330.

<sup>723</sup> Vgl. Berger et al. 2019, S. 20.

## Studie 2

RPI womöglich nochmals deutlich verstärkt werden. Sofern RPI eine zusätzliche Informationsfunktion ist, kann dies durch die Untersuchung weiterer, ergänzender Feedbackmechanismen validiert werden. Falls die Wirkung von RPI in der zusätzlichen Dimension des Feedbacks liegt, kann es durch andere Feedbackdimensionen substituiert werden. Das heißt, es können unterschiedliche zusätzliche Feedbackdimensionen in das Experiment eingeführt werden. Diese können sowohl in ihrer Art als auch in der Häufigkeit untersucht werden. Hier stellt sich die rudimentäre Frage, ob *mehr* Feedback *besseres* Feedback ist und ob die Art des Feedbacks nachrangig gesehen werden kann.

## 6 Erkenntnisfortschritt

Der Aufbau von Kapitel 6 wird in Abbildung 57 dargestellt. Es wird in Abschnitt 6.1 gezeigt, wie die Entwicklung und Definition eines Algorithm Aversion Bias durch diese Arbeit gestützt wird. In Abschnitt 6.2 werden zunächst die studienübergreifenden Limitationen aufgezeigt und es werden Implikationen für die künftige Forschung gegeben. Darauf aufbauend, werden die empirischen Erkenntnisse dieser Arbeit, um Implikationen für die Praxis erweitert. Diese basieren auf dem Austausch mit Praktikern. In Abschnitt 6.3 erfolgt eine abschließende Bewertung der empirischen Ergebnisse dieser Arbeit.

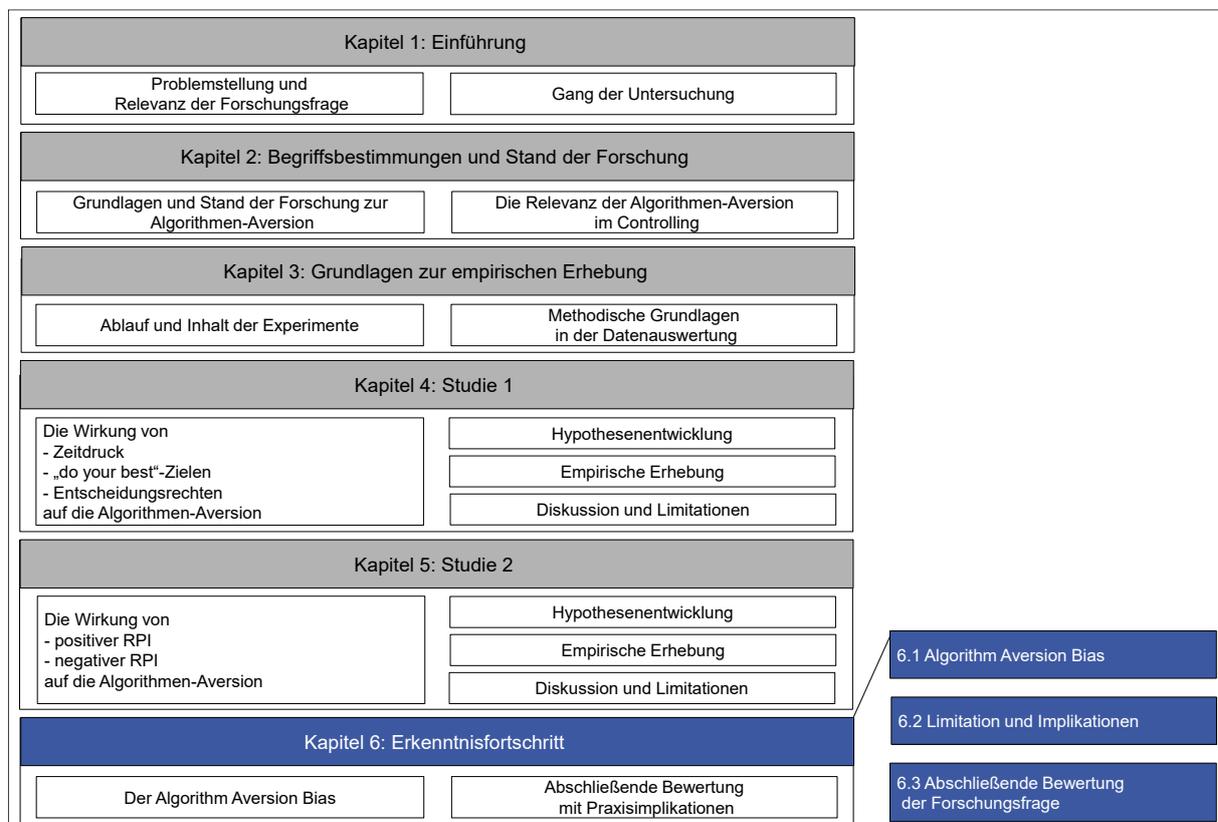


Abbildung 57: Aufbau von Kapitel 6<sup>724</sup>

### 6.1 Algorithm Aversion Bias

Die kausalen Zusammenhänge, die im Rahmen dieser Arbeit untersucht wurden, orientierten sich an den bestehenden Erkenntnissen der Algorithmen-Aversion. Da aufgrund der Neuartigkeit der Algorithmen-Aversion nicht von einer Theorie gesprochen

<sup>724</sup> Eigene Darstellung.

werden konnte,<sup>725</sup> wurden die zugrundeliegenden Annahmen im Sinne eines theoretischen Modells bzw. eines theoretischen Analyserahmens betrachtet. Eine Theorie benötigt ein abgeschlossenes und konsistentes Aussagensystem mit hohem Bewährungsgrad, das als noch nicht erreicht angesehen werden kann.<sup>726</sup> Um die Entwicklung hin zu einer Theorie der Algorithmen-Aversion zu unterstützen, wird in dieser Arbeit der Vorschlag eines *Algorithm Aversion Bias* gemacht.

Biases treten unter Unsicherheit auf und führen zu suboptimalen Entscheidungen, da Menschen die rationale Urteilsfindung systematisch missachten. In dieser Arbeit konnte dieses Verhalten für die Algorithmen-Aversion bestätigt werden. Zur Systematisierung des Algorithm Aversion Bias können die Ausführungen zu kognitiven Biases in Abschnitt 2.1.4.1 zusammengefasst definiert werden nach Arnott (2006):

„Cognitive biases are cognitions or mental behaviours that prejudice decision quality [a] in a significant number of decisions [b] for a significant number of people [c]; they are inherent in human reasoning. Cognitive biases are often called decision biases or judgement biases. One way of viewing cognitive biases is as predictable deviations from rationality. A rational choice is one based on the decision-maker's current assets and the possible consequences of the choice [d]“<sup>727</sup>.

Die Algorithmen-Aversion führt in allen identifizierten Arbeiten des derzeitigen Stands der Forschung und in den empirischen Untersuchungen dieser Arbeit, zu einer schlechten Forecastinggenauigkeit und damit zu signifikanten monetären Verlusten bei Forecastern [a]. Außerdem wurde sie im weiteren, wie auch im engeren Sinne, über unterschiedliche Aufgabenbereiche und damit in unterschiedlichen Entscheidungssituationen nachgewiesen [b].<sup>728</sup> Über alle Experimentalgruppen in dieser Untersuchung hinweg hat sich die Algorithmen-Aversion als äußerst robust erwiesen. Zwar konnte sie gemindert werden, jedoch war sie bei allen Untersuchungen gegeben. Auch die Literatur unterstreicht die Robustheit einer Algorithmen-Aversion [c].<sup>729</sup> In Konklusion aus diesen drei Argumenten und den Überlegungen zur Prinzipal-Agenten-Theorie in Abschnitt 2.1 geht hervor, dass die Ablehnung des Algorithmus keine rationale Entscheidung, sondern kognitiv verzerrt ist [d]. Damit die Algorithmen-Aversion nicht mehr nur als vereinzelt auftretender Effekt gesehen wird, sollte in Zukunft vom

---

<sup>725</sup> Siehe Abschnitt 1.2.

<sup>726</sup> Vgl. Döring und Bortz 2016, S. 59.

<sup>727</sup> Arnott 2006, S. 59.

<sup>728</sup> Vgl. Castelo et al. 2019, S. 809.

<sup>729</sup> Vgl. Filiz et al. 2021, o. S.; Logg et al. 2019, S. 90.

Algorithm Aversion Bias gesprochen werden. Die dem Algorithm Aversion Bias zugrunde liegende Heuristik, kann in der Fokussierung auf die Fehler des Algorithmus durch Leistungsfeedback gesehen werden. Bestätigt wird dies durch die Untersuchung von Zeitdruck und RPI. Wenn eine Defokussierung, weg von den Fehlern des Algorithmus und eine starke Fokussierung, durch Druck oder zusätzliches Feedback, auf die schlechtere eigene Leistung auftreten, kann diese Heuristik abgeschwächt und der Algorithm Aversion Bias verringert werden.

## 6.2 Limitationen und Implikationen

### 6.2.1 Studienübergreifende Limitationen und künftiger Forschungsbedarf

Die in Abschnitt 4.4.5 bzw. Abschnitt 5.4.4 gezeigten Limitationen und künftigen Forschungsbedarfe beziehen sich auf die untersuchten Hypothesen in Studie 1 bzw. Studie 2 und den damit verbundenen spezifischen Herausforderungen und Erkenntnisse. Eine abschließende Bewertung des Erkenntnisfortschritts erfordert darüber hinaus eine Diskussion der studienübergreifenden Limitationen. Diese lassen sich nach der Stichprobe und dem Experiment systematisieren. Tabelle 48 zeigt die wichtigsten Aspekte der Limitationen und des künftigen Forschungsbedarfs.

Die Auswahl der Stichprobe über Amazon MTurk folgt Dietvorst et al. (2015). Auch wenn zwischen den Erhebungen dieser Arbeit und Dietvorst et al. (2015) rund fünf Jahre liegen, kann nicht allumfänglich sichergestellt werden, dass die Teilnehmer dieser Studien nicht schon einmal an einem ähnlichen Experiment teilgenommen haben.<sup>730</sup> Begegnet wurde dieser Limitation durch eine Screening Frage zur Teilnahme an ähnlichen Studien in der Vergangenheit.<sup>731</sup> Weitere Einschränkungen die aus der Stichprobe resultieren, sind teilweise übertragbar auf das grundsätzliche Problem der externen Validität von Experimentalstudien. Die externe Validität wird der in Abschnitt 3.1.1 genannten internen Validität nachgeordnet. Sie ist „[...] das Ausmaß [...] in dem experimentell erzielte Befunde auf andere Gruppen, Umgebungen und Situationen übertragen werden können“<sup>732</sup>. Auch wenn die Studien dieser Arbeit durch MTurker in deren gewohntem und natürlichem Arbeitsumfeld durchgeführt wurden und damit im

---

<sup>730</sup> Zwischen Studie 1 und Studie 2 ist dies sichergestellt. Die Kenntnis über die MTurk-ID der Teilnehmer in Studie 1 ermöglicht einen Ausschluss dieser MTurker für Folgestudien.

<sup>731</sup> Siehe Abbildung 21.

<sup>732</sup> Eifler 2014, S. 205.

weiteren Sinne als Feldexperimente angesehen werden können, muss die externe Validität hinterfragt werden. Hinsichtlich der Stichprobe besteht daher eine Limitation im so genannten *Selection Bias*.<sup>733</sup>

Tabelle 48: Limitationen und künftiger Forschungsbedarf

	Stichprobe	Experiment
<b>Limitationen</b>	<ul style="list-style-type: none"> <li>- Selection Bias, MTurker können nicht alle typischen Merkmale von Forecastern im Controlling abdecken (Einsatz eines Algorithmus über Wochen oder Monate, geringes Expertenwissen in einer Aufgabe).</li> <li>- Fehlende Implikationen aus dem Alter und dem Bildungshintergrund von Forecastern</li> </ul>	<ul style="list-style-type: none"> <li>- Forecast des Algorithmus korreliert hoch mit der Bezahlregel.</li> <li>- Leistungsfeedback als reine Gegenüberstellung von Forecast mit wahren Wert.</li> <li>- Leistungsfeedback kann als Training gesehen werden, welches Einfluss auf das Verhalten gehabt haben kann.</li> </ul>
<b>Künftiger Forschungsbedarf</b>	<ul style="list-style-type: none"> <li>- Übertragung der Experimente in das Feld (zu Praktikern).</li> <li>- Untersuchung der Algorithmen-Aversion bei jüngeren und älteren Forecastern.</li> <li>- Untersuchung der Algorithmen-Aversion vor dem Bildungshintergrund von Forecastern bzw. vor und nach einer Schulung zu Algorithmen.</li> </ul>	<ul style="list-style-type: none"> <li>- Übertragung der Experimente in das Feld (zu Praxisanwendungen).</li> <li>- Untersuchung unterschiedlicher Algorithmen.</li> <li>- Untersuchung der Häufigkeit von Leistungsfeedback.</li> <li>- Erweiterung der Darstellung von Leistungsfeedback.</li> <li>- Untersuchungen zu mehrperiodigen Zusammenhängen (Target Ratcheting).</li> </ul>

Die Literatur zeigt, dass durch eine ansteigende Overconfidence, wie sie bei Forecastingexperten vorherrscht, die Algorithmen-Aversion höher ausfällt.<sup>734</sup> Weiter

<sup>733</sup> Vgl. Peel 2018, S. 173. Durch die Randomisierung in der Gruppenzuteilung wurde eine unverzerrte Verteilung der Stichproben sichergestellt.

<sup>734</sup> Siehe Logg et al. 2019, S. 90.

wurde auch gezeigt, dass durch die häufige Zusammenarbeit mit Algorithmen, die Algorithmen-Aversion geringer ausgeprägt ist.<sup>735</sup> Dies sind zwei Einflussfaktoren, die bedingt durch die Stichprobe, in der vorliegenden Untersuchung gering ausgeprägt sind. MTurker sind zwar de facto Forecaster die für einen Arbeitslohn an der Aufgabe die ihnen gegeben wird arbeiten, sie arbeiten jedoch weder wiederholt in dieser Aufgabe, noch wiederholt mit dem zur Verfügung stehenden Algorithmus. Ergo können sie nur bedingt den Status eines Experten für die zugrunde liegende Aufgabe erreichen. Durch die hohe Varianz der Aufgaben in MTurk gibt es keine Standardtools zur Lösung der Aufgaben und es ist anzunehmen, dass die MTurker in der Regel Aufgaben eigenständig, also ohne Hilfsmittel wie Algorithmen, bearbeiten. Ein im Experiment eingeführter Algorithmus ist demnach *neu* für MTurker. Die hohe Sensibilität für fehlerhaftes Verhalten des Algorithmus kann dadurch womöglich erklärt werden. Eine Reduktion der Algorithmen-Aversion durch die Zusammenarbeit mit einem Algorithmus über mehrere Aufgaben oder über einen längeren Zeitraum hinweg erfolgt dadurch ebenfalls nicht. Aus diesen Überlegungen heraus ist es notwendig die Forschung zur Algorithmen-Aversion in die Praxis zu tragen.

Eine Limitation, die aus der Stichprobe resultiert, liegt in den soziodemographischen Gegebenheiten. Dabei kann die Frage aufgeworfen werden, inwiefern das Alter der Teilnehmer, einen Einfluss auf das Verhalten zum Einsatz von Algorithmen hat. Junge Forecaster sind womöglich vertrauter im Umgang mit Algorithmen als ältere Forecaster. Die steigende Algorithmen-Appreciation in Abschnitt 4.4.2 legt hier nahe, dass mit jüngeren Forecastern die Algorithmen-Aversion noch weiter ansteigen kann. Es stellt sich jedoch die Frage, wie elastisch dieser Effekt ist. Das heißt, dass eine sehr starke Algorithmen-Appreciation in Zukunft nicht zu einer steigenden Algorithmen-Aversion führt sondern diese überstrahlt. Das heißt der Effekt der Algorithmen-Aversion verschwindet bei sehr starker Algorithmen-Appreciation.

Weiter gibt die Forschung bislang kaum Indikationen darüber, wie sich der Bildungshintergrund von Forecastern auf die Algorithmen-Aversion auswirkt. Logg et al. (2019) haben gemeinsam mit den Teilnehmern ihrer Experimente eine Definition eines Algorithmus entwickelt. Dabei sind sie zu dem Schluss gekommen, dass die Mehrheit

---

<sup>735</sup> Siehe Filiz et al. 2021, o. S.

der Teilnehmer ein einheitliches, technisch geprägtes, Verständnis von einem Algorithmus hat.<sup>736</sup> Womöglich entsteht durch diesen technischen Charakter eine verstärkte Algorithmen-Aversion. Durch das Schulen von logischem Denken und der Funktionsweise von Algorithmen kann eine höhere Akzeptanz für Algorithmen geschaffen werden.<sup>737</sup> Weiter können unterschiedliche Beschreibungen zu Algorithmen und deren Arbeitsweise bzw. Herkunft zur Senkung der Algorithmen-Aversion erforscht werden.

Die Forschung zur Algorithmen-Aversion beruht bislang im Wesentlichen auf Online-Experimenten.<sup>738</sup> Daher liegt eine bedeutende Limitation der bisherigen Forschung wiederholt darin, dass der Untersuchungskontext nicht in die Unternehmenspraxis erweitert wurde. Für Feldexperimente stellt der experimentelle Aufbau nur bedingt Einschränkungen dar und kann damit weitestgehend übertragen werden. Das heißt die Einführung von mehrperiodigen verwendeten Algorithmen mit Leistungsfeedback ist eng an die Praxis angelehnt und kann daher analog untersucht werden. Zur besseren Vergleichbarkeit der späteren Ergebnisse sollten möglichst ähnliche Aufgaben verwendet werden.

Der in dieser Untersuchung verwendete Algorithmus ist zu Replikationszwecken derselbe, wie er von Dietvorst et al. (2015) eingesetzt wurde. Dabei weist der Algorithmus eine Besonderheit auf. Die Genauigkeit des Algorithmus ist hochkorreliert mit der Bezahlregel. Das heißt das Leistungsspektrum des Algorithmus ist besonders vorteilhaft um relativ sicher ein hohes Incentive mit dem Algorithmus zu erreichen.<sup>739</sup> Forecaster, die diesen Zusammenhang in den Übungsforecasts wahrnehmen, werden unabhängig von weiteren Einflussfaktoren, eine Tendenz zugunsten des Algorithmus entwickeln.

Als Auslöser der Algorithmen-Aversion wurde das Leistungsfeedback definiert.<sup>740</sup> Leistungsfeedback beruht auf zehn Übungsforecasts. Die Wirkung von Leistungsfeedback wurde robust in Studie 1 bestätigt. Ohne Leistungsfeedback, also bei

---

<sup>736</sup> Vgl. Logg et al. 2019, S. 93 und Abschnitt 2.1.1.

<sup>737</sup> Vgl. Matt et al. 2016, S. 103; Nihalani et al. 2011, S. 776. Siehe hierzu auch die Ausführungen in Abschnitt 2.2.1 zum Berufsbild des Data Scientists.

<sup>738</sup> Siehe dazu beispielsweise Castelo et al. 2019, S. 809; Dietvorst et al. 2018, S. 1155; Logg et al. 2019, S. 90.

<sup>739</sup> Siehe Abschnitt 3.2.3.

<sup>740</sup> Siehe Abschnitt 2.1.3.

## Erkenntnisfortschritt

keiner Interaktion mit Algorithmen vor dem incentivierten Forecast, werden diese bereitwillig akzeptiert. Filiz et al. (2021) zeigen, dass durch Leistungsfeedback über 40 Forecasts die Algorithmen-Aversion sinkt.<sup>741</sup> Sie begründen dies durch den Lerneffekt, der in der Vielzahl an Forecasts entsteht. In der Studie von Feng und Gao (2020) wurde darüber hinaus gezeigt, dass bei 50 Interaktionen mit einem Algorithmus die Algorithmen-Aversion reduziert wird.<sup>742</sup> Begründet wird dies durch die Regret-Aversion, wenn der Vergleich mit einem Algorithmus häufig stattfindet und den Teilnehmern wiederholt vor Augen geführt wird, dass der Algorithmus eine bessere Leistung als sie selbst erbringen kann.<sup>743</sup> Bei der ex post demonstrierten falschen Wahl zum eigenen Forecast wird die getroffene Entscheidung vom Forecaster bereut. Wird die Algorithmen-Aversion demnach über die Anzahl der Interaktionen mit dem Algorithmus betrachtet entsteht das Bild einer inversen U-Kurve (siehe Abbildung 58). Wie beständig dieser Effekt über unterschiedliche Häufigkeiten des Leistungsfeedbacks ist und wie stark seine Auswirkungen auf die Algorithmen-Aversion tatsächlich sind sollte in künftigen Arbeiten untersucht werden.

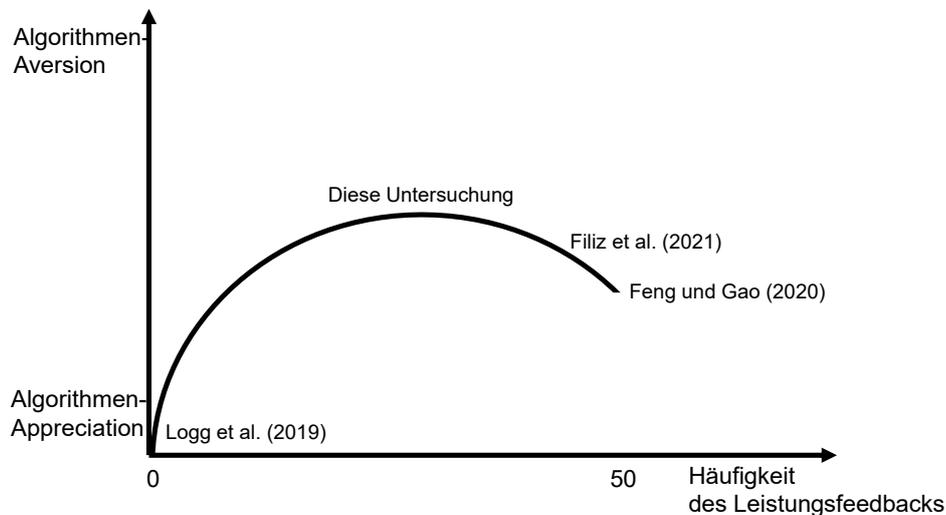


Abbildung 58: Verlauf der Algorithmen-Aversion über die Häufigkeit des Leistungsfeedbacks<sup>744</sup>

<sup>741</sup> Vgl. Filiz et al. 2021, o. S. sowie Abschnitt 2.1.5.11.

<sup>742</sup> Vgl. Feng und Gao 2020, o. S.

<sup>743</sup> Vgl. Feng und Gao 2020, o. S. sowie Abschnitt 2.1.5.7.

<sup>744</sup> Eigene Darstellung.

Ein weiterer Aspekt der in der künftigen Forschung adressiert werden sollte ist die Darstellung des Leistungsfeedbacks. Das Leistungsfeedback wurde den Teilnehmern in Form eines Vergleichs ihrer absoluten Einschätzung mit der des Algorithmus präsentiert. Das Feedback könnte künftig anders präsentiert werden, um den unerwünschten Feedback-Effekt abzumildern. Dem Forecaster können prozentuale Abweichungen pro Forecast und die mittlere Abweichung mehrerer Forecasts angezeigt werden. Dem Forecaster kann eine Leistungszusammenfassung gegeben werden, um das Feedback auf neutrale und objektive Weise darzustellen. So können Effekte vermieden werden, die aus einer Überreaktion aufgrund vermeintlich schlechter Leistungen des Algorithmus<sup>745</sup> oder einiger sehr guter menschlicher Forecasts resultieren.<sup>746</sup> Darüber hinaus kann die Präsentation von Feedback durch neue Technologien wie Sprachausgabe oder KPI-Dashboards optimiert werden.<sup>747</sup>

Die bisherigen Untersuchungen zur Algorithmen-Aversion unterscheiden häufig zwischen dem Fokus auf *einen* incentivierten Forecast oder *mehreren* incentivierten Forecasts. Im Falle mehrerer incentivierter Forecasts wurde gezeigt, dass ein lernfähiger Algorithmus die Aversion abbaut.<sup>748</sup> Das heißt hier stehen die sequenziellen Forecasts in einer gewissen Abhängigkeit zueinander, welche wiederum die Algorithmen-Aversion beeinflusst. Neben einem lernenden Algorithmus, kann die Abhängigkeit zwischen sequenziellen Forecasts auch in der Abhängigkeit der Incentivierung zwischen mehreren Perioden gesehen werden. Ziele sind in der Praxis häufig nicht statisch definiert, sondern dynamisch ausgerichtet, das heißt die Zielerreichung in einer Periode, beeinflusst das Ziel in der Folgeperiode. Dies wiederum kann zum sogenannten *Target Ratcheting* führen.<sup>749</sup> Im Target Ratcheting entsteht ein dynamisches Anreizproblem, bei dem Forecaster Anreize dazu erhalten ihre aktuelle Leistung zu drosseln, um ihre zukünftigen Ziele leichter zu erreichen.<sup>750</sup> Diese Manipulation zeigt sich beispielsweise darin, dass Vertriebs-Forecaster bei hinreichender Zielerreichung in einem bestimmten Jahr, im letzten Quartal dieses Jahres ihre Leistung reduzieren und Abschlüsse in das Folgejahr legen.<sup>751</sup> Mit Target Ratcheting geht damit in einem hohen Maße die

---

<sup>745</sup> Vgl. Petropoulos et al. 2016, S. 842; Remus et al. 1995, S. 285.

<sup>746</sup> Vgl. Choi und Hui 2014, S. 614; Grieco und Hogarth 2009, S. 756.

<sup>747</sup> Vgl. Bentley et al. 2018, S. 23.

<sup>748</sup> Vgl. Berger et al. 2021, S. 55.

<sup>749</sup> Vgl. Weitzman 1980, S. 302.

<sup>750</sup> Vgl. Bouwens und Kroos 2011, S. 171; Indjejikian et al. 2014, S. 1227.

<sup>751</sup> Vgl. Leone und Rock 2002, S. 43.

Notwendigkeit der Kontrolle über die Leistungserbringung einher. Durch die Steigerung der Ziele über die Zeit anhand vorhergehender Zielerreichung, werden Forecaster dazu motiviert, ihre Leistung entsprechend dem gesetzten Ziel optimal zu steuern.<sup>752</sup> Das heißt, es wird eine vollständige Zielerfüllung angestrebt, ohne dabei das Ziel zu übertreffen, selbst wenn ein Übertreffen problemlos möglich wäre. Wenn bestimmte Aspekte einer Aufgabe durch einen Algorithmus unterstützt bzw. ersetzt werden, wird gleichbeutend eine gewisse Kontrolle der Aufgabenerfüllung abgeben. Deshalb hat der Forecaster, wenn er einen Algorithmus einsetzt, weniger Möglichkeiten, die Leistung zu reduzieren und damit Potenziale zurückzuhalten. Eine theoretische Schlussfolgerung daraus ist, dass die Forecaster, die über Target Ratcheting gesteuert werden, eine stärkere Algorithmen-Aversion haben. Daraus ergibt sich, dass Unternehmen ihren Zielvereinbarungsprozess anpassen müssen, wenn sie die Algorithmen-Aversion der Forecaster senken wollen.<sup>753</sup> Dieser Zusammenhang wurde bislang noch nicht empirisch untersucht.

### 6.2.2 Praxisimplikationen

Dieser Abschnitt basiert auf den Erkenntnissen der empirischen Untersuchungen und dem Stand der Forschung sowie dem Dialog mit Praktikern. Dieser Dialog beinhaltet einen Austausch mit elf Controllern beim Stuttgarter Arbeitskreis des Internationalen Controller Vereins (ICV) und sechs Experteninterviews (siehe Anhang 5). Der Austausch und die Interviews fanden von Februar bis März 2022 statt. Die Interviews wurden mit prozessverantwortlichen in Forecasts und der Entwicklung und Einführung von Algorithmen in Forecasts durchgeführt.<sup>754</sup>

Die Forschungsfrage zielte darauf ab, wie sich eine Berücksichtigung des Arbeitsumfeldes auf die Einführung von Algorithmen auswirkt. Dabei wurden Variablen untersucht, die für das Arbeitsumfeld von Forecastern widerspiegeln. Für die betriebliche Praxis ist diese Interaktion vom Arbeitsumfeld mit Algorithmen ein komplexes Feld. Beherrscht wird diese Komplexität beispielsweise durch die organisatorische Gestaltung gemäß dem Business Analytics.<sup>755</sup> Hier werden eine *individuelle*-, *Team*- und *Unternehmensebene* betrachtet. Darüber hinaus findet sich eine Algorithmen-Aversion

---

<sup>752</sup> Vgl. Indjejikian et al. 2014, S. 1227.

<sup>753</sup> Vgl. Mitra et al. 1997, S. 117.

<sup>754</sup> Siehe Anhang 5.

<sup>755</sup> Für personelle und organisatorischen Rahmenbedingungen, die sich aus der Einführung von Business Analytics ergeben, siehe Seiter 2019, S. 66–104.

in gesellschaftlichen bzw. *politischen Aspekten* wieder. Entsprechend werden die Implikationen in diesem Abschnitt gegliedert nach einer individuellen-, Team- und Unternehmensebene, sowie der politischen Ebene:

Auf individueller Ebene zeigt die Praxis, dass Forecaster tiefergehend in den Prozess der Einführung von Algorithmen involviert werden, als dies in dieser Untersuchung umgesetzt wurde. Das heißt, wichtige Algorithmen werden in Kooperation mehrerer Mitarbeiter wie Data Scientists, Informatiker oder Mathematiker eingeführt.<sup>756</sup> Insbesondere der Bezug und das Commitment von Controllern zu Prozessen und Aufgaben, sind in der Praxis deutlich ausgeprägter als in experimentellen Studien bisher gezeigt wurde. Der Forecaster, der mit dem jeweiligen Algorithmus zusammenarbeitet wird zum sogenannten *Process-Owner*.<sup>757</sup> Das heißt, er ist nicht nur verantwortlich für die Nutzung sondern, auch für die Einführung des Algorithmus. Er muss sich daher bereits vor der Nutzung mit dessen Gütekriterien befassen und kann dabei aktiv den Algorithmus gestalten und verbessern. Eine alternative Rolle zum Process-Owner stellt der *Data and Model Interpreter* dar.<sup>758</sup> Die Ausbildung hin zu einem Data and Model Interpreter erweitern die Rolle des Process-Owner um spezifische Aufgaben und Fähigkeiten: „People who fill this role develop visualizations and/or models and understand and embrace the fact that part of their job is to communicate and explain the BI/analytics output“<sup>759</sup>. Dadurch sollen die Forecasts nicht nur visuell optimiert gestaltet werden, sondern auch durch eine geeignete Geschichte (sogenanntes *Storytelling*) in überzeugender Weise präsentiert werden.<sup>760</sup> Damit soll eine Identifikation des Forecasters mit den eingeführten Algorithmen erzeugt werden, welche die Aversion gegenüber diesen zu einem Widerspruch mit der eigenen Leistung erwachsen lässt. Darüber hinaus werden in den Interviews insbesondere Potenziale in der Schulung und Weiterbildung gesehen.<sup>761</sup> Schulung und Weiterbildung in statistischen Aspekten ist dabei von besonderer Relevanz. Es sollen nicht nur der Forecast und der Algorithmus *richtig* interpretiert werden, sondern vielmehr soll die Fähigkeit ausgebildet werden, bei Unklarheiten die richtigen Fragen zu stellen. Die Fähigkeit, gezielte Fragen zu

---

<sup>756</sup> Vgl. Interview E und F.

<sup>757</sup> Vgl. Kokina und Blanchette 2019, S. 6 und Interview D.

<sup>758</sup> Vgl. Watson 2017, S. 164.

<sup>759</sup> Watson 2017, S. 164.

<sup>760</sup> Vgl. Interview A.

<sup>761</sup> Vgl. Interview B und D.

unverständlichen Aspekten stellen zu können, kann ein effektives Instrument zur Reduktion der Algorithmen-Aversion darstellen. Es kann dadurch einerseits ein Overconfidence Bias erzeugt werden, denn die Forecaster werden befähigt den Algorithmus zu besitzen bzw. zu beherrschen. Andererseits, wird durch das gesteigerte Verständnis und Commitment zum Prozess und zum Algorithmus eine hohe Akzeptanz erwartet. Der Kontext der Entscheidungsrechte über die Gestaltung und damit verbunden über die Eingabeinformationen des Algorithmus, nimmt demnach in der Praxis eine immer zentralere Rolle ein. Dabei ist vor dem Hintergrund der empirischen Ergebnisse dieser Arbeit jedoch zu hinterfragen, ob dies zu einer dauerhaften Reduktion der Algorithmen-Aversion führt.

Aus Sicht der Teamebene stellen die Implikationen aus Studie 2 die Grundlagen dar. Es wurde postuliert, dass RPI dann eingesetzt werden soll, wenn Algorithmen etwas besser sind als die Forecaster, die mit ihnen arbeiten.<sup>762</sup> Forecaster mit guten Leistungen unterstützen diese Algorithmen und verbessern sie, Forecaster die schlechter als ein Algorithmus sind, können wiederum vom Algorithmus profitieren. Die Forecaster mit guter Leistung werden also bevorzugt zu Process-Ownern und Data and Model Interpreter. Dabei ist anzumerken, dass für umfangreiche Aufgaben und Algorithmen mehrere Forecaster, mit unterschiedlichen Fähigkeiten, sich diese Rollen teilen können.<sup>763</sup> Unter der Prämisse, dass Eingriffe von Forecastern in das Ergebnis des Algorithmus den Forecast in der Regel schlechter machen, sollten diese grundsätzlich minimiert werden. Jedoch sollten sie zugelassen werden, wenn es starke Eingriffe sind und wenn sie die Akzeptanz für einen Algorithmus erhöhen.<sup>764</sup> Aus Sicht der Praxis wird hier auf einen spielerischen Umgang mit Algorithmen verwiesen. Ein Algorithmus, der nicht durch einen Forecaster angelernt wird, sondern durch ein ganzes Team, kann als Mitglied dieses Teams verstanden werden. Ein solcher spielerischer Umgang mit Algorithmen kann die Aversion absenken.<sup>765</sup> Neben der Betrachtung von Fähigkeiten innerhalb eines Teams, sind zwischen der Teamebene und der Unternehmensebene der Vorgesetzte und dessen Delegationsbefugnisse zu nennen.<sup>766</sup> Der

---

<sup>762</sup> Siehe Abschnitt 5.4.3.

<sup>763</sup> Vgl. Watson 2017, S. 164.

<sup>764</sup> Vgl. Fildes et al. 2009, S. 3.

<sup>765</sup> Vgl. Interview B. Oshri und Plugge, S. 151 berichten von der Bezeichnung von RPA-Bots als Komikhelden: „[T]he development team gave each bot a comic hero name, and insisted on referring to the bot by its name in any communications“.

<sup>766</sup> Vgl. Interview C und Nagar 2002, S. 379.

Einsatz von Algorithmen wurde im Kontext dieser Arbeit als freiwillige Entscheidung gesehen, welche zumeist durch den Einsatz von Incentives für den bestmöglichen Forecast gesteuert wird. Die Interviewpartner äußern jedoch Kritik an dieser Sichtweise.<sup>767</sup> Die Einführung von Algorithmen wird häufig vom Vorgesetzten als direkte Anweisung durchgesetzt. Das heißt, dass der Einsatz von Algorithmen eine explizite Arbeitsanweisung ist und entsprechend keine optionale Arbeitsstrategie. Dies wiederum sichert den weitläufigen Einsatz von Algorithmen ab und nimmt gleichzeitig die Verantwortung vom einzelnen Forecaster.

Darüber hinaus ist bei der Gestaltung von Teams zu berücksichtigen, dass durch den zukünftig verstärkten Einsatz von Homeoffice, innerhalb von Teams, eine größere soziale Distanz entsteht.<sup>768</sup> Das eröffnet die Frage, inwiefern Expertenmeinungen, innerhalb von Teams, durch eine soziale Distanz, eine Abwertung erhalten. Die höhere Gewichtung einer Expertenmeinung gegenüber einem Algorithmus kann daher zukünftig zugunsten von Algorithmen verlagert werden und die Algorithmen-Aversion gesenkt werden. Dazu trägt die womöglich sinkende Confidence in den Experten bei, da der persönliche Austausch mit diesem geringer wird.<sup>769</sup>

Auf der Unternehmensebene ist zu berücksichtigen, dass nicht nur die Forecaster, die direkt an einer Aufgabe beteiligt sind das Ergebnis eines Algorithmus beurteilen, sondern auch das Management oder externe Stakeholder beteiligt sind.<sup>770</sup> Diese werden nur geringfügig in die Prozesse zur Entstehung des Forecasts oder des Algorithmus eingebunden. Dadurch sind die vorgelagerten Mechanismen wie sie bei Process-Ownern in der Entwicklung des Algorithmus entstehen nicht relevant. Es ist mit Hinblick auf die starke Wirkung des Leistungsfeedback der Zeitpunkt für die Einführung von Algorithmen dadurch entscheidend und mit Bedacht zu wählen. Das heißt, Algorithmen sollten erst dann eingeführt werden, wenn sie zuverlässige Ergebnisse erzeugen. Auch wenn bereits sichergestellt werden kann, dass die Ergebnisse eines Algorithmus die eines Menschen übertreffen, sollte die Einführung erst dann erfolgen, wenn eine hohe Zuverlässigkeit des Algorithmus sichergestellt werden kann. Das bedeutet

---

<sup>767</sup> Vgl. Interview A und C.

<sup>768</sup> Vgl. Interview B.

<sup>769</sup> Vgl. auch Jussupow et al. 2020, S. 10.

<sup>770</sup> Vgl. Interview D.

darüber hinaus, dass der Einführungsprozess die Berücksichtigung externer Störfaktoren benötigt.<sup>771</sup> Wenn starke interne oder externe Veränderungen absehbar sind, werden sowohl die Forecasts des Algorithmus als auch die des Forecasters schlechter. Da dies jedoch nur dem Algorithmus negativ ausgelegt wird, sollten in turbulenten Zeiten, weniger neue Algorithmen eingeführt werden – es sei denn, es besteht ein hoher Zeitdruck, denn dieser kann die Algorithmen-Aversion auf einem niedrigen Niveau halten.

Die Implikationen auf politischer Ebene erfolgen am Beispiel der Forecasts zur Entwicklung der Coronapandemie. Die deutsche Bevölkerung wollte im Zuge des Coronavirus, trotz Warnungen aus Wissenschaft und Politik, nur widerwillig Selbstbeschränkungen zum Wohle der Allgemeinheit hinnehmen. Dabei kann die Frage aufgeworfen werden, weshalb ein Großteil der Bevölkerung sein Verhalten so Widerwillig an den Empfehlungen der Politik ausgerichtet hat.

Als Mitte Januar 2020 erstmals ausführlich über das Coronavirus in China berichtet wurde, wurde das Risiko einer umfangreichen Ausbreitung zunächst als gering beschrieben.<sup>772</sup> Häufig wurde ein Vergleich zum SARS-Erreger Anfang der 2000er Jahre gezogen, um die Kontrollierbarkeit einer Corona-Infektion zu bekräftigen. Es handle sich um eine Form der Grippe, welche lediglich bei sehr engem Kontakt mit Erkrankten übertragbar sei und sich dadurch nur langsam verbreite.<sup>773</sup> Nachdem das Virus allerdings in Europa angekommen war, erlangte es durch schnell steigende Fallzahlen sowie dem Versagen des Gesundheitssystems in Italien, rasch die ungeteilte Aufmerksamkeit der deutschen Bevölkerung und Politik. Dieses Interesse an der Entwicklung wiederum wurde von wissenschaftlichen Forecasts begleitet. Trotz der Aussage des Robert Koch Instituts, dass sich auf längere Sicht bis zu 60 Millionen Deutsche infizieren könnten und deshalb soziale Kontakte gemieden werden sollten, zeigte sich ein Großteil der Bevölkerung kaum bereit, seine sozialen Aktivitäten freiwillig einzuschränken.<sup>774</sup>

---

<sup>771</sup> Vgl. Interview G und D.

<sup>772</sup> Vgl. Niederer 2020, o. S.

<sup>773</sup> Vgl. Niederer 2020, o. S.

<sup>774</sup> Vgl. Tagesschau 2020, o. S.

Dafür können mehrere Begründungen herangezogen werden. Zunächst spielen Informationen in Form von wissenschaftlichen Erkenntnissen, welche an die Öffentlichkeit getragen werden, eine essenzielle Rolle. Das Problem dabei ist, dass seit Januar 2020 wiederholt Aussagen revidiert und Erkenntnisse bezüglich Ansteckungsgefahr und Risikogruppe überholt wurden – es gab damit ein Leistungsfeedback zu den getätigten Forecasts.<sup>775</sup> Grund dafür war die Neuartigkeit der Erkrankung und der Druck auf Wissenschaft und Forschung, Antworten bzw. Forecasts zur Entwicklung zu liefern. Die kontinuierliche und zeitnahe Versorgung der Bevölkerung mit neusten Informationen aus der Forschung führte dazu, dass Schlüsse zu schnell gezogen und aufgrund fehlender Erfahrungswerte ungenaue Forecasts getroffen wurden. Durch diese Unsicherheit, welche Fakten dauerhaft richtig sind und welche es nicht sind, wird die Glaubwürdigkeit wissenschaftlicher Erkenntnisse und in diesem Fall der Prognosemodelle, von der Bevölkerung infrage gestellt und es kann mit einer aufkommenden Algorithmen-Aversion argumentiert werden. Dies kann dazu führen, dass Forecasts stark hinterfragt werden und dadurch keine Verhaltensanpassung stattfindet. Im Falle der Corona-Prognosen war die Verlässlichkeit der Prognosedaten ein zentrales Problem. Die Datengrundlage war aufgrund der geringen Erfahrung mit dem Virus vergleichsweise schlecht. Zudem beziehen unterschiedliche Prognosemodelle unterschiedliche Faktoren in ihre Analyse mit ein, woraus abweichende Forecasts zwischen den Modellen unterschiedlicher Institutionen entstehen. Beispielsweise basieren Prognosen auf unterschiedlichen Übertragungsraten. Obwohl es einen großen Unterschied macht, ob eine Person statistisch gesehen eine oder vier Personen ansteckt, findet sich hierbei kein einheitliches Vorgehen.<sup>776</sup>

Eine hohe Qualität von Forecasts wurde insofern erschwert, dass offizielle Zahlen denen der tatsächlichen Entwicklung hinterherhinken. Ob diese Vorhersage dann den tatsächlichen Verlauf widerspiegelt oder lediglich auf der Höhe der Testkapazitäten der Labore basiert, ist außerdem fraglich.<sup>777</sup> Trotz vieler solcher Unsicherheiten sehen Wissenschaftler einen großen Nutzen in vagen Prognosemodellen. Hier titulierte der Deutschlandfunk den Wissenschaftsjournalist Volkart Wildermuth: „Warum selbst vage Prognosen nützlich sein können“<sup>778</sup>. Aus Sicht der Algorithmen-Aversion ist hier

---

<sup>775</sup> Vgl. Gasteiger 2020, o. S.

<sup>776</sup> Vgl. Heiden und Buchholz 2020, S. 7.

<sup>777</sup> Vgl. Wildermuth und Blumenthal 2020, o. S.

<sup>778</sup> Wildermuth und Blumenthal 2020, o. S.

eindeutig eine Synchronisierung der Prognosemodelle notwendig und vage Modelle sollten vermieden werden. Ein Wettbewerb zwischen Institutionen um die genauesten Forecasts wird durch die Publikation vieler schlechter Prognosemodelle zu einer hohen Aversion gegenüber den Prognosemodellen führen.

Die genannten Aspekte sind lediglich eine Auswahl derer Faktoren, welche einen möglichen Einfluss auf das Prognoseergebnis haben können. Durch das uneinheitliche Vorgehen der Wissenschaft wird der Bevölkerung eine Fülle an Informationen bereitgestellt, die sich entweder nach einiger Zeit als falsch herausstellen oder koexistieren. Damit ist erklärbar, dass die Confidence in diese Modelle sinkt und Menschen versuchen, sich ihre persönliche Meinung aus den vorhandenen Daten bestätigen zu lassen. Das heißt die quantitativen Modelle werden durch anekdotischen Evidenz aus Einzelfällen überlagert. Diese Einzelfälle wiederum unterliegen keinem Fehler, sondern sind direkt überprüfbar – sie sind jedoch nicht repräsentativ. Es entsteht eine subjektive Überlegenheit von Einzelfällen gegenüber quantitativen Forecasts, welche wiederum die Algorithmen-Aversion bestärken.

Das kann anschließend zu unterschiedlichen Verhaltensweisen der Menschen führen: Die einen erkennen die Gefahr nicht an und die anderen tätigen Hamsterkäufe. Das heißt Menschen die Forecasts keinen Glauben schenken, erkennen den Ernst der Lage nicht, da sie sich persönlich keinem höheren Risiko ausgesetzt fühlen. Die aus den Prognosemodellen erhaltene Evidenz, durch das eigene Verhalten andere, – möglicherweise gefährdetere Personen – zu infizieren, wird nicht angenommen. Bezüglich der Hamsterkäufe zeigt sich ein anderes Extrem. Zwar existieren keine seriösen Prognosen, die geschlossene Lebensmittelläden und Schwierigkeiten in der Warenversorgung voraussagen, dennoch decken sich Menschen mit Dingen wie Desinfektionsmittel oder Masken ein, die bspw. Kranke oder Pflegende dringender bräuchten und nicht zum Überleben gesunder, im Haus Isolierter beitragen.<sup>779</sup> Dieses Beispiel kann als eine Wirkung des Zeitdrucks interpretiert werden. Auch wenn nicht allen Empfehlungen der Politik gefolgt wird, führt ein aufkommender Zeitdruck zur Beschaffung von Lebensmitteln, was mit einer sinkenden Algorithmen-Aversion gegenüber den Prognosemodelle selbst interpretiert werden kann.

---

<sup>779</sup> Vgl. Ludwig 2020, o. S.

Eine weitere Besonderheit der Prognosen zum Coronavirus ist die Tatsache, dass es sich nicht um eine lineare, sondern vielmehr eine exponentielle Entwicklung der Verbreitung handelt.<sup>780</sup> Beim exponentiellen Wachstum findet jeweils eine Verdopplung der Fallzahl in einem bestimmten Zeitraum statt. Diese Entwicklung sendet im Anfangsstadium falsche Signale, da sich die Verbreitungskurve nur langsam steigert. Mit fortschreitender Verbreitung wird diese allerdings immer steiler und die Fallzahlen steigen sehr stark. Dies führt zu einem gewissen Zeitdruck in der Entscheidungsfindung bei Eindämmungsmaßnahmen. Mit Hinblick auf die Impfung bzw. die Impfquote kann interpretiert werden, dass die schlechten Prognosen zwar der Politik die Überzeugungskraft nehmen, jedoch durch den exponentiell wachsenden Zeitdruck ein gewisser Handlungsdruck in der Bevölkerung erzeugt wurde. Neben Privilegien für Geimpfte gegenüber Ungeimpften, wurde durch den Zeitdruck auf die Eindämmung des Virus eine relativ hohe Impfquote ermöglicht. Konkret wirkte sich dies so aus, dass die Impfquote in Deutschland Ende 2021 bei rund 74 % lag, während jeder Zweite unzufrieden mit der selbigen war. Das heißt die Algorithmen-Aversion wurde auch hier durch Zeitdruck abgesenkt.

### **6.3 Abschließende Bewertung der Forschungsfrage**

Ziel der vorliegenden Arbeit war es, ein besseres Verständnis für die die Algorithmen-Aversion in Forecasts zu entwickeln. Da die Algorithmen-Aversion ein robustes Phänomen ist, hat die Forschung ein großes Interesse daran entwickelt sie besser zu verstehen.<sup>781</sup> Die Literatur hat hierbei bislang das Arbeitsumfeld von Forecastern vernachlässigt. Dies ist jedoch von großer Relevanz, da die tägliche Arbeit von Forecastern in diesem erfolgt. Das heißt, damit die Algorithmen-Aversion in Zukunft besser verstanden, aktiv gesteuert und langfristig gesenkt werden kann, ist es bedeutend zu verstehen, wie stark sie in Abhängigkeit des Arbeitsumfeldes auftritt. Die hier identifizierte Forschungslücke wurde zusammengefasst mit:

#### **Wie wirkt sich das Arbeitsumfeld von Forecastern auf die Algorithmen-Aversion aus?**

Um das Arbeitsumfeld von Forecastern hinsichtlich der Algorithmen-Aversion systematisch zu untersuchen, wurde für diese Arbeit auf das Framework von Bonner

---

<sup>780</sup> Vgl. Endt et al. 2020, o. S.

<sup>781</sup> Vgl. unter anderem Berger et al. 2021; Castelo et al. 2019; Dietvorst et al. 2018; Efendić et al. 2020; Filiz et al. 2021; Logg et al. 2019.

und Sprinkle (2002) aufgebaut.<sup>782</sup> Dafür wurden Variablen aus dem Arbeitsumfeld von Controllern identifiziert, die von hoher Relevanz für die Erstellung von Forecasts sind. Die für die vorliegende Arbeit relevanten Variablen waren: Zeitdruck, „do your best“-Ziele, Entscheidungsrechte, für Eingabeinformationen die der Algorithmus verarbeitet, und relatives Leistungsfeedback zu Kollegen (RPI). Tabelle 49 fasst die untersuchten Hypothesen und deren erwartete Wirkung sowie empirische Beurteilung zusammen.

Tabelle 49: Übersicht aller Hypothesentests dieser Arbeit

Hypothese	Prognostizierte Wirkung	Hypothesentest
H1: Durch Leistungsfeedback entsteht eine Algorithmen-Aversion.	↑	<input checked="" type="checkbox"/>
H2: Zeitdruck reduziert die Algorithmen-Aversion.	↓	<input checked="" type="checkbox"/>
H3: Ein „do your best“-Ziel reduziert die Algorithmen-Aversion.	↓	<input type="checkbox"/>
H4: Entscheidungsrechte für Eingabeinformationen, die der Algorithmus verarbeitet, reduzieren die Algorithmen-Aversion.	↓	<input type="checkbox"/>
H5: Positive RPI zur Forecastinggenauigkeit erhöht die Algorithmen-Aversion.	↑	<input checked="" type="checkbox"/>
H6: Negative RPI zur Forecastinggenauigkeit reduziert bei leistungsschwächeren Forecastern die Algorithmen-Aversion.	↓	<input checked="" type="checkbox"/>

Mittels H1 wurde die Grundlage für diese Untersuchung geschaffen und es wurde gezeigt, dass durch Leistungsfeedback eine Algorithmen-Aversion vorherrscht. Daraus konnten zwei Schlüsse gezogen werden. Einerseits liegt bei Forecastern eine starke Algorithmen-Appreciation vor und andererseits ist damit eine ansteigende Algorithmen-Aversion verbunden. Durch die regelmäßige Interaktion mit Algorithmen im privaten wie auch beruflichen Umfeld entsteht eine Routine in der Zusammenarbeit mit Algorithmen.<sup>783</sup> Dies führt zu einer Algorithmen-Appreciation, bei der Algorithmen bereitwillig angenommen werden, obwohl deren Leistung und Präzision nicht bekannt ist. Ein Vergleich mit den Daten aus Dietvorst et al. (2015) zeigt darüber hinaus, dass diese Algorithmen-Appreciation über die Zeit anzusteigen scheint. Dies ist insofern zu

<sup>782</sup> Vgl. Bonner und Sprinkle 2002, S. 303.

<sup>783</sup> Vgl. Griffin und Wright 2015, S. 377; Quattrone 2016, S. 118.

hinterfragen, da bei unzureichendem Wissen über die Struktur und Zuverlässigkeit eines Algorithmus keine Aussagen über die Qualität des Algorithmus getroffen werden können und ihm somit eine kritische Skepsis entgegengebracht werden sollte.<sup>784</sup> In der empirischen Untersuchung führt die Algorithmen-Aversion zu einer geringeren Genauigkeit der Forecasts und resultiert damit in einer geringeren leistungsabhängigen Vergütung für die Forecaster. Es liegt somit eine für Forecaster und darauf aufbauen für Unternehmen eine wertvernichtende Algorithmen-Aversion vor.

Mit H2 konnte gezeigt werden, dass die Algorithmen-Aversion abgeschwächt wird, wenn Forecaster bei der Erstellung eines Forecasts, unter Zeitdruck stehen. Dies ist damit zu begründen, dass unter Zeitdruck die Confidence in den eigenen Forecast sinkt. Mit dieser Reduktion der Confidence in den eigenen Forecast geht eine geringere Algorithmen-Aversion einher. Es wird daher von einer Mediation der sinkenden Algorithmen-Aversion durch eine sinkende Confidence in den eigenen Forecast gesprochen. Da Zeitdruck in der bestehenden Literatur als etwas Negatives beschrieben wird, da es Stress auslöst und häufig zu schlechter Leistung führt, wurde hier ein wichtiger Beitrag zum wissenschaftlichen, ebenso wie zum praktischen Erkenntnisfortschritt geleistet.<sup>785</sup> Zeitdruck ist basierend auf den empirischen Resultaten dieser Arbeit nicht als per se schlecht anzusehen,<sup>786</sup> insbesondere dann nicht, wenn einem Forecaster ein Algorithmus zur Verfügung steht, der in der Drucksituation zur Unterstützung eingesetzt werden kann. Es steht dadurch eine Erleichterung der täglichen Arbeit bei gleichzeitiger Verbesserung der Arbeitsergebnisse zur Verfügung. Zeitdruck ist außerdem eine Erweiterung der Möglichkeiten, wie ein Algorithmus in Forecasts verstärkt eingesetzt wird, ohne dass ein Mensch in den Forecast eines Algorithmus eingreift und damit die Forecastinggenauigkeit des Algorithmus verschlechtert.<sup>787</sup> Dies kann aktiv gesteuert werden, indem Daten vom Management zeitverzögert bereitgestellt werden oder die verfügbare Zeit für die Erstellung von Forecasts durch die Verkürzung von Fristen verknappt wird. Da die Algorithmen-Aversion unter Zeitdruck reduziert wird, jedoch in Teilen noch besteht, darf der Zeitdruck nur in moderatem Ausmaß erzeugt werden.<sup>788</sup>

---

<sup>784</sup> Vgl. Fildes et al. 2009, S. 3; McKinney et al. 2017, S. 63.

<sup>785</sup> Vgl. Pietsch und Messier 2017, S. 51; Wegier und Spaniol 2015.

<sup>786</sup> Vgl. Kirchler et al. 2017, S. 57.

<sup>787</sup> Vgl. Dietvorst et al. 2018, S. 1155; Fildes et al. 2009, S. 3.

<sup>788</sup> Vgl. Kirchler et al. 2017, S. 57.

In H3 wurde ein „do your best“-Ziel für die Forecastinggenauigkeit eingeführt. Dieses ersetzt das in der Forschung zur Algorithmen-Aversion etablierte schwierige und spezifische Ziel eines möglichst perfekten Forecasts. Negative Effekte von schwierigen und spezifischen Zielen können eine enge Fokussierung auf die Zielerreichung durch die intensive Arbeit mit bestehenden Methoden erzeugen.<sup>789</sup> Durch das „do your best“-Ziel sollte es Forecastern ermöglicht werden, die überlegene, aber nicht perfekte Leistung des Algorithmus besser einzuordnen und ihn entsprechend öfter zu nutzen. Die empirischen Ergebnisse bestätigen H3 nicht. Im Wesentlichen hat sich keine Änderung im Verhalten durch das „do your best“-Ziel gezeigt. Dazu muss ergänzt werden, dass die mit dem „do your best“-Ziel einhergehende Incentivierung den Forecastern einen größeren Bonus garantiert hat. Zusammengefasst bedeutet das, dass die Forecastingergebnisse nicht verbessert werden, die Algorithmen-Aversion weiterhin besteht und die Incentivierung für diese schlechten Forecasts besonders hoch ausfällt.

Mit H4 wurde in Bezug auf die Entscheidungsrechte von Forecastern die Hypothese aufgestellt, dass durch die Möglichkeit der Einflussnahme auf den Informationsinput des Algorithmus, letzterer öfter eingesetzt wird und die Algorithmen-Aversion sinkt. Die Literatur zeigte dazu bereits im Vorfeld, dass die Algorithmen-Aversion reduziert werden kann, indem Forecaster die Forecastingergebnisse eines Algorithmus verändern dürfen.<sup>790</sup> Die Anpassung des Forecasts des Algorithmus führt zu einer Verschlechterung der Forecastinggenauigkeit, weshalb solche Eingriffe zumeist nicht sinnvoll sind.<sup>791</sup> Durch die Einführung von Entscheidungsrechten für die Eingabeinformationen zeigen die empirischen Ergebnisse keine Unterschiede im Verhalten der Forecaster und dementsprechend auch keine Veränderung in der Algorithmen-Aversion. Es wird angenommen, dass die Forecaster im Rahmen dieser Arbeit kein hinreichendes Verständnis über die Datengrundlage erlangt haben. Sie fühlten sich demnach nicht befähigt, eine Auswahl an Eingabeinformationen zu treffen die den Forecast des Algorithmus verbessert. In Bezug auf die Entscheidungsrechte bei der Arbeit mit Algorithmen muss daher festgehalten werden, dass die Algorithmen-Aversion nur dann

---

<sup>789</sup> Vgl. Gold et al. 2014, S. 303; Seckler et al. 2017, S. 22.

<sup>790</sup> Vgl. Dietvorst et al. 2018, S. 1155.

<sup>791</sup> Siehe dazu beispielsweise Dietvorst et al. 2018, S. 1155; Goodwin und Fildes 1999, S. 37.

verringert werden kann, wenn die Forecaster das Ergebnis eines Algorithmus beeinflussen können – nicht aber die Eingabeinformationen.

Durch H5 und H6 wurde die Auswirkung von relativem Leistungsfeedback zu Kollegen, sogenanntes RPI untersucht. RPI ergänzt das bestehende Leistungsfeedback, welches die Algorithmen-Aversion auslöst, um eine weitere Feedbackdimension. Zur effektiven Untersuchung der Wirkung von RPI wurde in negative und positive RPI unterschieden. Darüber hinaus wurde für negative RPI unterschieden, wie gut die tatsächlich Leistung eines Forecasters war. Es wurde sowohl für positive wie auch negative RPI gezeigt, dass eine signifikante Wirkung auf die Algorithmen-Aversion besteht. Das heißt bei positiver RPI steigt die Algorithmen-Aversion. Bei negativer RPI und schlechter Performance hingegen, sinkt die Algorithmen-Aversion.

Liegt positive RPI vor, wurden Forecaster darin bestärkt, dass sie eine gute Leistung erbringen – ihre Algorithmen-Aversion steigt dadurch. Da auch die Leistung dieser Forecaster schlechter als die des Algorithmus war, ist positive RPI im Kontext der Algorithmen Aversion weitestgehend zu vermeiden.<sup>792</sup> Negative RPI kann Forecaster, die relativ schlechte Forecasts erstellen, dabei unterstützen, vermehrt Algorithmen zu wählen und dadurch ihre Forecastinggenauigkeit zu verbessern. Diese Forecaster sind bei nur Feedback zu ihrer schlechten Leistung unzureichend in der Lage ihre schlechte Leistung im Vergleich zum besseren Algorithmus richtig einzuschätzen. Das negative RPI zeigt ihnen auf, dass sie im Vergleich mit ihren Peers ebenfalls keine gute Leistung erbringen. So wie in der Betrachtung zur Algorithmen-Aversion unter Zeitdruck in H2, unterstützt RPI den Ausweg aus schlechter Leistung und erwachsenem Druck, der nicht aus eigenem Handeln heraus reduziert werden kann.<sup>793</sup>

Die empirischen Ergebnisse dieser Arbeit zeigen, dass das Arbeitsumfeld von Forecastern einen wesentlichen Einfluss auf deren Algorithmen-Aversion haben kann. Damit erweitert die vorliegende Arbeit die Literatur zur Algorithmen-Aversion um:

- die Möglichkeiten zur Senkung der Algorithmen-Aversion durch Zeitdruck und negative RPI bei schlechter Performance.
- die Risiken einer steigenden Algorithmen-Aversion durch positive RPI.

---

<sup>792</sup> Potenziell sinnvolle Strategien zum Einsatz von RPI im Kontext der Algorithmen-Aversion finden sich in Abschnitt 5.4.3.

<sup>793</sup> Es muss hier angemerkt werden, dass die RPI in der vorliegenden Arbeit keinen Zusammenhang zur Vergütung hatte. Diese war auch unter RPI an die Forecastinggenauigkeit im incentivierten Forecast gekoppelt.

## Erkenntnisfortschritt

- das Wissen darüber, dass eine Senkung insbesondere dann eintritt, wenn die eigene Leistung realistischer eingeschätzt werden kann und ein Algorithmus dadurch als Verbesserung von Forecasts gesehen wird.

Darüber hinaus bestätigt die vorliegende Arbeit:

- die hohe Relevanz der Algorithmen-Aversion, da Forecaster auch bei schlechterer Leistung als ein Algorithmus diesen nicht einsetzen und sie durch die Algorithmen-Aversion eine geringere Vergütung realisieren.
- die Robustheit der Algorithmen-Aversion, da sie gegenüber unterschiedlichen Gegebenheiten des Arbeitsumfelds wie „do your best“-Zielen oder Entscheidungsrechten für Eingabeinformationen zum Algorithmus bestand hat.

## 7 Literaturverzeichnis

- Akin, Z.; Karagözoğlu, E. (2017): The Role of Goals and Feedback in Incentivizing Performance. In: *Managerial and Decision Economics* 38 (2), S. 193–211. DOI: 10.1002/mde.2753.
- Alder, G. S. (2007): Examining the relationship between feedback and performance in a monitored environment: A clarification and extension of feedback intervention theory. In: *The Journal of High Technology Management Research* 17 (2), S. 157–174. DOI: 10.1016/j.hitech.2006.11.004.
- Alder, G. S.; Ambrose, M. L. (2005): Towards understanding fairness judgments associated with computer performance monitoring: An integration of the feedback, justice, and monitoring research. In: *Human Resource Management Review* 15 (1), S. 43–67. DOI: 10.1016/j.hrmr.2005.01.001.
- Al-Htaybat, K.; Alberti-Alhtaybat, L. von (2017): Big Data and corporate reporting: impacts and paradoxes. In: *Accounting, Auditing & Accountability Journal* 30 (4), S. 850–873. DOI: 10.1108/AAAJ-07-2015-2139.
- Alicke, M. D.; Govorun, O. (2005): The better-than-average effect. In: *The Self in Social Judgment* 1, S. 85–106.
- Amir, E.; Ganzach, Y. (1998): Overreaction and underreaction in analysts' forecasts. In: *Journal of Economic Behavior & Organization* 37 (3), S. 333–347. DOI: 10.1016/S0167-2681(98)00092-4.
- Anthony, R. N. (1965): Planning and control systems. A framework for analysis. Boston: Harvard Univ (Studies in management control).
- Appelbaum, D.; Kogan, A.; Vasarhelyi, M.; Yan, Z. (2017): Impact of business analytics and enterprise systems on managerial accounting. In: *International Journal of Accounting Information Systems* 25, S. 29–44. DOI: 10.1016/j.accinf.2017.03.003.
- Arnold, M. C.; Artz, M. (2015): Target difficulty, target flexibility, and firm performance: Evidence from business units' targets. In: *Accounting, Organizations and Society* 40, S. 61–77. DOI: 10.1016/j.aos.2014.12.002.

- Arnold, M. C.; Hannan, R. L.; Tafkov, I. D. (2018): Team Member Subjective Communication in Homogeneous and Heterogeneous Teams. In: *The Accounting Review* 93 (5), S. 1–22. DOI: 10.2308/accr-52002.
- Arnott, D. (2006): Cognitive biases and decision support systems development: a design science approach. In: *Information Systems Journal* 16 (1), S. 55–78. DOI: 10.1111/j.1365-2575.2006.00208.x.
- Ashraf, N.; Bandiera, O.; Lee, S. S. (2014): Awards unbundled: Evidence from a natural field experiment. In: *Journal of Economic Behavior & Organization* 100, S. 44–63. DOI: 10.1016/j.jebo.2014.01.001.
- Ashton, R. H. (1990): Pressure and Performance in Accounting Decision Settings: Paradoxical Effects of Incentives, Feedback, and Justification. In: *Journal of Accounting Research* 28, S. 148. DOI: 10.2307/2491253.
- Azmat, G.; Iriberry, N. (2010): The importance of relative performance feedback information: Evidence from a natural experiment using high school students. In: *Journal of Public Economics* 94 (7-8), S. 435–452. DOI: 10.1016/j.jpubeco.2010.04.001.
- Backhaus, K.; Erichson, B.; Gensler, S.; Weiber, R.; Weiber, T. (2021): *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung.* 16., vollständig überarbeitete und erweiterte Auflage. Wiesbaden: Springer Gabler.
- Baiman, S. (1990): Agency research in managerial accounting: A second look. In: *Accounting, Organizations and Society* 15 (4), S. 341–371. DOI: 10.1016/0361-3682(90)90023-N.
- Bandiera, O.; Barankay, I.; Rasul, I. (2013): Team Incentives: Evidence from a Firm Level Experiment. In: *Journal of the European Economic Association* 11 (5), S. 1079–1114. DOI: 10.1111/jeea.12028.
- Baron, R. M.; Kenny, D. A. (1986): The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. In: *Journal of Personality and Social Psychology* 51 (6), S. 1173–1182. DOI: 10.1037//0022-3514.51.6.1173.

- Bartneck, C.; Duenser, A.; Moltchanova, E.; Zawieska, K. (2015): Comparing the similarity of responses received from studies in Amazon's Mechanical Turk to studies conducted online and with direct recruitment. In: *PLoS one* 10 (4), e0121595. DOI: 10.1371/journal.pone.0121595.
- Baur, N.; Blasius, J. (Hg.) (2014): Handbuch Methoden der empirischen Sozialforschung. Wiesbaden: Springer VS (Handbuch).
- Becker, W.; Brandt, B. (2016): Anreizsysteme im Controlling. In: Wolfgang Becker und Patrick Ulrich (Hg.): Handbuch Controlling. Wiesbaden, Heidelberg: Springer Gabler (Springer NachschlageWissen), S. 459–473.
- Beilock, S. L.; Carr, T. H. (2001): On the fragility of skilled performance: What governs choking under pressure? In: *Journal of Experimental Psychology: General* 130 (4), S. 701–725. DOI: 10.1037/0096-3445.130.4.701.
- Ben Zur, H.; Breznitz, S. J. (1981): The effect of time pressure on risky choice behavior. In: *Acta Psychologica* 47 (2), S. 89–104. DOI: 10.1016/0001-6918(81)90001-9.
- Bénabou, R.; Tirole, J. (2002): Self-confidence and personal motivation. In: *The Quarterly Journal of Economics* 117 (3), S. 871–915.
- Ben-David, I.; Graham, J. R.; Harvey, C. R. (2013): Managerial Miscalibration. In: *The Quarterly Journal of Economics* 128 (4), S. 1547–1584. DOI: 10.1093/qje/qjt023.
- Bentley, F.; Luvogt, C.; Silverman, M.; Wirasinghe, R.; White, B.; Lottridge, D. (2018): Understanding the Long-Term Use of Smart Speaker Assistants. In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2 (3), S. 1–24. DOI: 10.1145/3264901.
- Berger, B.; Adam, M.; Rühr, A.; Benlian, A. (2021): Watch Me Improve-Algorithm Aversion and Demonstrating the Ability to Learn. In: *Business & Information Systems Engineering* 63 (1), S. 55–68. DOI: 10.1007/s12599-020-00678-5.
- Berger, L.; Fiolleau, K.; MacTavish, C. (2019): I Know Something You Don't Know: The Effect of Relative Performance Information and Individual Performance Incentives on Knowledge Sharing. In: *Journal of Management Accounting Research* 31 (2), S. 19–35. DOI: 10.2308/jmar-52152.
- Bergmann, M.; Brück, C.; Knauer, T.; Schwering, A. (2020): Digitization of the budgeting process: determinants of the use of business analytics and its effect on satisfaction

- with the budgeting process. In: *Journal of Management Control* 25 (3), S. 29. DOI: 10.1007/s00187-019-00291-y.
- Berinsky, A. J.; Huber, G. A.; Lenz, G. S. (2012): Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk. In: *Political Analysis* 20 (3), S. 351–368. DOI: 10.1093/pan/mpr057.
- Bisbe, J.; Malagueño, R. (2012): Using strategic performance measurement systems for strategy formulation: Does it work in dynamic environments? In: *Management Accounting Research* 23 (4), S. 296–311. DOI: 10.1016/j.mar.2012.05.002.
- Blanes i Vidal, J.; Nossol, M. (2011): Tournaments without prizes: Evidence from personnel records. In: *Management Science* 57 (10), S. 1721–1736.
- Bohn, U. (2007): Vertrauen in Organisationen: Welchen Einfluss haben Reorganisationsmaßnahmen auf Vertrauensprozesse? Dissertation. Unter Mitarbeit von Stefan Kühl.
- Bondt, W. de; Thaler, R. (1985): Does the Stock Market Overreact? In: *The Journal of Finance* 40 (3), S. 793–805. DOI: 10.1111/j.1540-6261.1985.tb05004.x.
- Bonner, S. E. (1999): Judgment and Decision-Making Research in Accounting. In: *Accounting Horizons* 13 (4), S. 385–398. DOI: 10.2308/acch.1999.13.4.385.
- Bonner, S. E. (2008): Judgment and decision making in accounting. 1. ed. Upper Saddle River, NJ: Pearson/Prentice Hall.
- Bonner, S. E.; Lewis, B. L. (1990): Determinants of Auditor Expertise. In: *Journal of Accounting Research* 28, S. 1. DOI: 10.2307/2491243.
- Bonner, S. E.; Sprinkle, G. B. (2002): The effects of monetary incentives on effort and task performance: theories, evidence, and a framework for research. In: *Accounting, Organizations and Society* 27 (4-5), S. 303–345. DOI: 10.1016/S0361-3682(01)00052-6.
- Bouwens, J.; Kroos, P. (2011): Target ratcheting and effort reduction. In: *Journal of Accounting and Economics* 51 (1-2), S. 171–185. DOI: 10.1016/j.jacceco.2010.07.002.
- Büchtemann, C. F.; Schupp, J.; Soloff, D. J. (1993): Übergänge von der Schule in den Beruf–Deutschland und USA im Vergleich. In: *Mitteilungen aus der Arbeitsmarkt-und Berufsforschung* 26 (4), S. 507–520.

Burton, J. W.; Stein, M.-K.; Jensen, T. B. (2019): A systematic review of algorithm aversion in augmented decision making. In: *Journal of Behavioral Decision Making* 27 (11), S. 1–20. DOI: 10.1002/bdm.2155.

Buschbacher, F.; Weber, M. (2017): Künstliche Intelligenz. Wirtschaftliche Bedeutung, gesellschaftliche Herausforderungen, menschliche Verantwortung. Hg. v. Bitkom e.V. Berlin.

Busenitz, L. W.; Barney, J. B. (1997): Differences between entrepreneurs and managers in large organizations: Biases and heuristics in strategic decision-making. In: *Journal of Business Venturing* 12 (1), S. 9–30. DOI: 10.1016/S0883-9026(96)00003-1.

Butler, S. A.; Ghosh, D. (2015): Individual differences in managerial accounting judgments and decision making. In: *The British Accounting Review* 47 (1), S. 33–45. DOI: 10.1016/j.bar.2014.09.002.

Camerer, C. (1981): General conditions for the success of bootstrapping models. In: *Organizational Behavior and Human Performance* 27 (3), S. 411–422. DOI: 10.1016/0030-5073(81)90031-3.

Camerer, C.; Ho, T.-H.; Chong, J.-K. (2004): A Cognitive Hierarchy Model of Games. In: *The Quarterly Journal of Economics* 119 (3), S. 861–898. DOI: 10.1162/0033553041502225.

Carbone, R.; Andersen, A.; Corriveau, Y.; Corson, P. P. (1983): Comparing for Different Time Series Methods the Value of Technical Expertise Individualized Analysis, and Judgmental Adjustment. In: *Management Science* 29 (5), S. 559–566. DOI: 10.1287/mnsc.29.5.559.

Carroll, E.; Marginson, D. (2021): Relative performance information and social comparisons: Exploring managers' cognitive, emotional and dysfunctional behavioral processes. In: *Management Accounting Research* 53, S. 100768. DOI: 10.1016/j.mar.2021.100768.

Carver, C. S.; Scheier, M. F. (1981): The self-attention-induced feedback loop and social facilitation. In: *Journal of Experimental Social Psychology* 17 (6), S. 545–568.

- Casler, K.; Bickel, L.; Hackett, E. (2013): Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. In: *Computers in Human Behavior* 29 (6), S. 2156–2160. DOI: 10.1016/j.chb.2013.05.009.
- Castelo, N.; Bos, M. W.; Lehmann, D. R. (2019): Task-Dependent Algorithm Aversion. In: *Journal of Marketing Research* 56 (5), S. 809–825. DOI: 10.1177/0022243719851788.
- Cavanaugh, M. A.; Boswell, W. R.; Roehling, M. V.; Boudreau, J. W. (2000): An empirical examination of self-reported work stress among U.S. managers. In: *Journal of Applied Psychology* 85 (1), S. 65–74. DOI: 10.1037/0021-9010.85.1.65.
- Chamoni, P.; Gluchowski, P. (2017): Business Analytics — State of the Art. In: *Controlling & Management Review* 61 (4), S. 8–17.
- Chen, C.; Keith, J. (2005): Are Companies Really Ready for Stretch Targets? In: *Management Accounting Quarterly* 6 (4), S. 10–18.
- Chen, G.; Crossland, C.; Luo, S. (2015): Making the same mistake all over again: CEO overconfidence and corporate resistance to corrective feedback. In: *Strategic Management Journal* 36 (10), S. 1513–1535. DOI: 10.1002/smj.2291.
- Chenhall, R. H. (2003): Management control systems design within its organizational context. Findings from contingency-based research and directions for the future. In: *Accounting, Organizations and Society* 28 (2-3), S. 127–168. DOI: 10.1016/S0361-3682(01)00027-7.
- Chew, N.; Houghton, K. A. (1996): Prediction of corporate failure: The effect of feedback on human judgments. In: *Asian Review of Accounting* 4 (1), S. 3–24. DOI: 10.1108/eb060663.
- Child, J.; Möllering, G. (2003): Contextual Confidence and Active Trust Development in the Chinese Business Environment. In: *Organization Science* 14 (1), S. 69–80. DOI: 10.1287/orsc.14.1.69.12813.
- Choi, D.; Hui, S. K. (2014): The role of surprise: Understanding overreaction and underreaction to unanticipated events using in-play soccer betting market. In: *Journal of Economic Behavior & Organization* 107, S. 614–629. DOI: 10.1016/j.jebo.2014.02.009.

- Christ, M. H.; Emett, S. A.; Tayler, W.; Wood, D. (2016): Compensation or feedback: Motivating performance in multidimensional tasks. In: *Accounting, Organizations and Society* 50, S. 27–40. DOI: 10.1016/j.aos.2016.03.003.
- Chuttur, M. (2009): Overview of the Technology Acceptance Model: Origins, Developments and Future Directions. Indiana University. Online verfügbar unter <http://sprouts.aisnet.org/9-37>.
- Cohen, J. (1983): The Cost of Dichotomization. In: *Applied Psychological Measurement* 7 (3), S. 249–253. DOI: 10.1177/014662168300700301.
- Cohen, J. (2013): *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hoboken: Taylor and Francis.
- Collins, K. M.; Killough, L. N. (1992): An empirical examination of stress in public accounting. In: *Accounting, Organizations and Society* 17 (6), S. 535–547. DOI: 10.1016/0361-3682(92)90012-H.
- Commerford, B.; Denny, S.; Joe, J.; Ulla, J. (2022): Man Versus Machine: Complex Estimates and Auditor Reliance on Artificial Intelligence. In: *Journal of Accounting Research* 60 (1), S. 171–201. DOI: 10.1111/1475-679X.12407.
- Corgnet, B.; Gómez-Miñambres, J.; Hernán-González, R. (2015): Goal Setting and Monetary Incentives: When Large Stakes Are Not Enough. In: *Management Science* 61 (12), S. 2926–2944. DOI: 10.1287/mnsc.2014.2068.
- Cornelissen, T.; Dustmann, C.; Schönberg, U. (2017): Peer Effects in the Workplace. In: *American Economic Review* 107 (2), S. 425–456. DOI: 10.1257/aer.20141300.
- Crump, M. J. C.; McDonnell, J. V.; Gureckis, T. M. (2013): Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. In: *PloS one* 8 (3), e57410. DOI: 10.1371/journal.pone.0057410.
- Crusius, J.; Mussweiler, T. (2012): To achieve or not to achieve? Comparative mindsets elicit assimilation and contrast in goal priming. In: *European Journal of Social Psychology* 42 (6), S. 780–788. DOI: 10.1002/ejsp.873.
- Dai, H.; Dietvorst, B. J.; Tuckfield, B.; Milkman, K. L.; Schweitzer, M. E. (2018): Quitting When the Going Gets Tough: A Downside of High Performance Expectations. In: *Academy of Management Journal* 61 (5), S. 1667–1691. DOI: 10.5465/amj.2014.1045.

- Das, T. K.; Teng, B.-S. (1999): Cognitive Biases and Strategic Decision Processes: An Integrative Perspective. In: *Journal of Management Studies* 36 (6), S. 757–778. DOI: 10.1111/1467-6486.00157.
- Davidson, R.; MacKinnon, J. G. (1993): Estimation and inference in econometrics. New York, NY, Oxford: Oxford University Press.
- Davis, F. D. (1985): A technology acceptance model for empirically testing new end-user information systems: Theory and results: Massachusetts Institute of Technology.
- Dawes, R.; Faust, D.; Meehl, P. (1989): Clinical versus actuarial judgment. In: *Science* 243 (4899), S. 1668–1674. DOI: 10.1126/science.2648573.
- Dawes, R. M. (1979): The robust beauty of improper linear models in decision making. In: *American Psychologist* 34 (7), S. 571–582. DOI: 10.1037/0003-066X.34.7.571.
- Dawes, R. M.; Corrigan, B. (1974): Linear models in decision making. In: *Psychological Bulletin* 81 (2), S. 95–106. DOI: 10.1037/h0037613.
- Delfgaauw, J.; Dur, R.; Non, A.; Verbeke, W. (2014): Dynamic incentive effects of relative performance pay: A field experiment. In: *Labour Economics* 28, S. 1–13. DOI: 10.1016/j.labeco.2014.02.003.
- Deloitte (2018): Forecasting in a digital world. Crunch time 6. Hg. v. Deloitte.
- DeNisi, A. S.; Kluger, A. N. (2000): Feedback effectiveness: Can 360-degree appraisals be improved? In: *Academy of Management Perspectives* 14 (1), S. 129–139.
- Deutsch, M. (1958): Trust and suspicion. In: *Journal of Conflict Resolution* 2 (4), S. 265–279. DOI: 10.1177/002200275800200401.
- Die Bundesregierung (2018): Strategie Künstliche Intelligenz der Bundesregierung. Nationale Strategie für Künstliche Intelligenz "AI Made in Germany".
- Dietvorst, B. J.; Bharti, S. (2020): People Reject Algorithms in Uncertain Decision Domains Because They Have Diminishing Sensitivity to Forecasting Error. In: *Psychological Science* 31 (10), S. 1302–1314. DOI: 10.1177/0956797620948841.
- Dietvorst, B. J.; Simmons, J. P.; Massey, C. (2015): Algorithm aversion. People erroneously avoid algorithms after seeing them err. In: *Journal of experimental psychology. General* 144 (1), S. 114–126. DOI: 10.1037/xge0000033.

- Dietvorst, B. J.; Simmons, J. P.; Massey, C. (2018): Overcoming Algorithm Aversion. People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them. In: *Management Science* 64 (3), S. 1155–1170. DOI: 10.1287/mnsc.2016.2643.
- Döring, N.; Bortz, J. (2016): Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften. 5. vollständig überarbeitete, aktualisierte und erweiterte Auflage. Berlin u.a.: Springer.
- Dror, I. E.; Busemeyer, J. R.; Basola, B. (1999): Decision making under time pressure: an independent test of sequential sampling models. In: *Memory & Cognition* 27 (4), S. 713–725. DOI: 10.3758/bf03211564.
- Earley, P. C.; Connolly, T.; & Ekegren, G. (1989): Goals, Strategy Development, and Task Performance: Some Limits on the Efficacy of Goal Setting. In: *Journal of Applied Psychology* 74 (1), S. 24–33.
- Edland, A.; Svenson, O. (1993): Time pressure and stress in human judgment and decision making. 1. ed. New York, NY: Plenum Press; Springer Science+Business Media.
- Efendić, E.; van de Calseyde, P. P.; Evans, A. M. (2020): Slow response times undermine trust in algorithmic (but not human) predictions. In: *Organizational Behavior and Human Decision Processes* 157, S. 103–114. DOI: 10.1016/j.obhdp.2020.01.008.
- Eifler, S. (2014): Experiment. In: Nina Baur und Jörg Blasius (Hg.): Handbuch Methoden der empirischen Sozialforschung. Wiesbaden: Springer VS (Handbuch), S. 195–211.
- Einhorn, H. J. (1986): Accepting error to make less error. In: *Journal of Personality Assessment* 50 (3), S. 387–395. DOI: 10.1207/s15327752jpa5003\_8.
- Eisenhardt, K. M. (1989): Agency Theory: An Assessment and Review. In: *The Academy of Management Review* 14 (1), S. 57. DOI: 10.2307/258191.
- Endt, C.; Mainka, M.; Müller-Hansen, S. (2020): Die Wucht der großen Zahl. Noch gibt es in Deutschland recht wenige Infizierte, doch das kann sich schnell ändern. Warum es wichtig ist, die Ausbreitung des Virus zu bremsen. In: *Süddeutsche Zeitung*, 10.03.2020. Online verfügbar unter <https://www.sueddeutsche.de/projekte/artikel/wissen/coronavirus-die-wucht-der-grossen-zahl-e575082/>, zuletzt geprüft am 31.05.2022.

- Eriksson, T.; Poulsen, A.; Villeval, M. C. (2009): Feedback and incentives: Experimental evidence. In: *Labour Economics* 16 (6), S. 679–688. DOI: 10.1016/j.labeco.2009.08.006.
- Eroglu, C.; Croxton, K. L. (2010): Biases in judgmental adjustments of statistical forecasts: The role of individual differences. In: *International Journal of Forecasting* 26 (1), S. 116–133. DOI: 10.1016/j.ijforecast.2009.02.005.
- Evanschitzky, H.; Armstrong, J. S. (2010): Replications of forecasting research. In: *International Journal of Forecasting* 26 (1), S. 4–8. DOI: 10.1016/j.ijforecast.2009.09.003.
- Eyring, H.; Ferguson, P.; Koppers, S. (2021): Less Information, More Comparison, and Better Performance: Evidence from a Field Experiment. In: *Journal of Accounting Research* 59 (2), S. 657–711. DOI: 10.1111/1475-679X.12362.
- Falk, A.; Ichino, A. (2006): Clean Evidence on Peer Effects. In: *Journal of Labor Economics* 24 (1), S. 39–57. DOI: 10.1086/497818.
- Faragher, J. (2021): Performance: Why ‘rank and yank’ fell out of favour. Online verfügbar unter <https://www.personneltoday.com/hr/why-rank-and-yank-fell-out-of-favour/>, zuletzt geprüft am 31.05.2022.
- Feichter, C.; Grabner, I. (2020): Empirische Forschung zu Management Control – Ein Überblick und neue Trends. In: *Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung* 72 (2), S. 149–181. DOI: 10.1007/s41471-020-00092-3.
- Feichter, C.; Grabner, I.; Moers, F. (2018): Target Setting in Multi-divisional Firms. State of the Art and Avenues for Future Research. In: *Journal of Management Accounting Research*. DOI: 10.2308/jmar-52158.
- Feng, X.; Gao, J. (2020): Is optimal recommendation the best? A laboratory investigation under the newsvendor problem. In: *Decision Support Systems* 131, S. 113251. DOI: 10.1016/j.dss.2020.113251.
- Festinger, L. (1954): A theory of social comparison processes. In: *Human Relations* 7 (2), S. 117–140.
- Fildes, R.; Beard, C. (1992): Forecasting Systems for Production and Inventory Control. In: *International Journal of Operations & Production Management* 12 (5), S. 4–27. DOI: 10.1108/01443579210011381.

- Fildes, R.; Goodwin, P.; Lawrence, M.; Nikolopoulos, K. (2009): Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning. In: *International Journal of Forecasting* 25 (1), S. 3–23. DOI: 10.1016/j.ijforecast.2008.11.010.
- Filiz, I.; Judek, J. R.; Lorenz, M.; Spiwoks, M. (2021): Reducing algorithm aversion through experience. In: *Journal of Behavioral and Experimental Finance* 31, S. 100524. DOI: 10.1016/j.jbef.2021.100524.
- Fleischman, G. M.; Johnson, E. N.; Walker, K. B. (2017): An Exploratory Examination of Management Accounting Service and Information Quality. In: *Journal of Management Accounting Research* 29 (2), S. 11–31. DOI: 10.2308/jmar-51614.
- Fong, C. J.; Patall, E. A.; Vasquez, A. C.; Stautberg, S. (2019): A Meta-Analysis of Negative Feedback on Intrinsic Motivation. In: *Educational Psychology Review* 31 (1), S. 121–162. DOI: 10.1007/s10648-018-9446-6.
- Frick, W. (2015): Here's Why People Trust Human Judgment Over Algorithms. Boston. Online verfügbar unter <https://hbr.org/2015/02/heres-why-people-trust-human-judgment-over-algorithms>.
- Galanti, S.; Vaubourg, A. G. (2017): Optimism bias in financial analysts' earnings forecasts: Do commissions sharing agreements reduce conflicts of interest? In: *Economic Modelling* 67, S. 325–337. DOI: 10.1016/j.econmod.2017.02.001.
- Garbers, Y.; Konradt, U. (2014): The effect of financial incentives on performance: A quantitative review of individual and team-based financial incentives. In: *Journal of Occupational and Organizational Psychology* 87 (1), S. 102–137. DOI: 10.1111/joop.12039.
- Gasteiger, C. (2020): Der Corona-Aufklärer der Nation. Virologe Christian Drosten. In: *Süddeutsche Zeitung*, 13.03.2020. Online verfügbar unter <https://www.sueddeutsche.de/medien/corona-drosten-virologe-1.4843374>, zuletzt geprüft am 31.05.2022.
- Germann, M.; Merkle, C. (2019): Algorithm Aversion in Financial Investing. In: *SSRN Journal*. DOI: 10.2139/ssrn.3364850.
- Gibbons, F. X.; Lane, D. J.; Gerrard, M.; Reis-Bergan, M.; Lautrup, C. L.; Pexa, N. A.; Blanton, H. (2002): Comparison-level preferences after performance: Is downward

comparison theory still useful? In: *Journal of Personality and Social Psychology* 83 (4), S. 865–880. DOI: 10.1037/0022-3514.83.4.865.

Gjedrem, W. G.; Kvaløy, O. (2020): Relative performance feedback to teams. In: *Labour Economics* 66, S. 101865. DOI: 10.1016/j.labeco.2020.101865.

Gluchowski, P. (2016): Business Analytics – Grundlagen, Methoden und Einsatzpotenziale. In: *HMD Praxis der Wirtschaftsinformatik* 53 (3), S. 273–286. DOI: 10.1365/s40702-015-0206-5.

Gold, A.; Gronewold, U.; Salterio, S. E. (2014): Error Management in Audit Firms: Error Climate, Type, and Originator. In: *The Accounting Review* 89 (1), S. 303–330. DOI: 10.2308/accr-50592.

Golden, J.; Milewicz, J.; Herbig, P. (1994): Forecasting: Trials and Tribulations. In: *Management Decision* 32 (1), S. 33–36.

Goodman, J. K.; Cryder, C. E.; Cheema, A. (2013): Data Collection in a Flat World: The Strengths and Weaknesses of Mechanical Turk Samples. In: *Journal of Behavioral Decision Making* 26 (3), S. 213–224. DOI: 10.1002/bdm.1753.

Goodwin, P.; Fildes, R. (1999): Judgmental forecasts of time series affected by special events: does providing a statistical forecast improve accuracy? In: *Journal of Behavioral Decision Making* 12 (1), S. 37–53. DOI: 10.1002/(SICI)1099-0771(199903)12:1<37::AID-BDM319>3.0.CO;2-8.

Gopalakrishnan, M.; Libby, T.; Samuels, J. A.; Swenson, D. (2015): The effect of cost goal specificity and new product development process on cost reduction performance. In: *Accounting, Organizations and Society* 42, S. 1–11. DOI: 10.1016/j.aos.2015.01.003.

Grieco, D.; Hogarth, R. M. (2009): Overconfidence in absolute and relative performance: The regression hypothesis and Bayesian updating. In: *Journal of Economic Psychology* 30 (5), S. 756–771. DOI: 10.1016/j.joep.2009.06.007.

Griffin, P. A.; Wright, A. M. (2015): Commentaries on Big Data's Importance for Accounting and Auditing. In: *Accounting Horizons* 29 (2), S. 377–379. DOI: 10.2308/acch-51066.

Gronau, N.; Thim, C.; Fohrholz, C. (2016): Business Analytics in der deutschen Praxis. In: *Controlling* 28 (8-9), S. 472–479. DOI: 10.15358/0935-0381-2016-8-9-472.

- Grove, W. M.; Zald, D. H.; Lebow, B. S.; Snitz, B. E.; Nelson, C. (2000): Clinical versus mechanical prediction: A meta-analysis. In: *Psychological Assessment* 12 (1), S. 19–30. DOI: 10.1037//1040-3590.12.1.19.
- Günther, T.; Boerner, X.; Mischer, M. (2021): Der Einfluss von Analytics Tools auf das Controlling: Erste Ergebnisse. DOI: 10.25368/2021.89.
- Guryan, J.; Kroft, K.; Notowidigdo, M. J. (2009): Peer Effects in the Workplace: Evidence from Random Groupings in Professional Golf Tournaments. In: *American Economic Journal: Applied Economics* 1 (4), S. 34–68. DOI: 10.1257/app.1.4.34.
- Häder, M. (2015): Empirische Sozialforschung. Eine Einführung. 3. Auflage. Wiesbaden: Springer.
- Hannan, R. L.; Krishnan, R.; Newman, A. H. (2008): The Effects of Disseminating Relative Performance Feedback in Tournament and Individual Performance Compensation Plans. In: *The Accounting Review* 83 (4), S. 893–913. DOI: 10.2308/accr.2008.83.4.893.
- Hannan, R. L.; McPhee, G. P.; Newman, A. H.; Tafkov, I. D. (2013): The Effect of Relative Performance Information on Performance and Effort Allocation in a Multi-Task Environment. In: *The Accounting Review* 88 (2), S. 553–575. DOI: 10.2308/accr-50312.
- Harrell, E. (2016): Managers Shouldn't Fear Algorithm-Based Decision Making. In: *Harvard Business Review*.
- Harvey, N. (1997): Confidence in judgment. In: *Trends in Cognitive Sciences* 1 (2), S. 78–82. DOI: 10.1016/S1364-6613(97)01014-0.
- Hattie, J.; Timperley, H. (2007): The Power of Feedback. In: *Review of Educational Research* 77 (1), S. 81–112. DOI: 10.3102/003465430298487.
- Hayes, A. F. (2009): Beyond Baron and Kenny: Statistical Mediation Analysis in the New Millennium. In: *Communication Monographs* 76 (4), S. 408–420. DOI: 10.1080/03637750903310360.
- Hayes, A. F. (2018): Introduction to mediation, moderation, and conditional process analysis. A regression-based approach. Second edition. New York, London: The Guilford Press (Methodology in the social sciences).

Hayes, A. F.; Cai, L. (2007): Using heteroskedasticity-consistent standard error estimators in OLS regression: an introduction and software implementation. In: *Behavior Research Methods* 39 (4), S. 709–722. DOI: 10.3758/BF03192961.

Hayes, A. F.; Rockwood, N. J. (2017): Regression-based statistical mediation and moderation analysis in clinical research: Observations, recommendations, and implementation. In: *Behaviour Research and Therapy* 98, S. 39–57. DOI: 10.1016/j.brat.2016.11.001.

Heiden, M. a. d.; Buchholz, U. (2020): Modellierung von Beispielszenarien der SARS-CoV-2-Epidemie 2020 in Deutschland. Robert Koch-Institut. Online verfügbar unter [https://www.rki.de/DE/Content/InfAZ/N/Neuartiges\\_Coronavirus/Modellierung\\_Deutschland.pdf?\\_\\_blob=publicationFile](https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Modellierung_Deutschland.pdf?__blob=publicationFile), zuletzt geprüft am 31.05.2022.

Herrmann, D. R.; Hope, O.-K.; Thomas, W. B. (2008): International Diversification and Forecast Optimism: The Effects of Reg FD. In: *Accounting Horizons* 22 (2), S. 179–197. DOI: 10.2308/acch.2008.22.2.179.

Heupel, T.; Lange, V. W.: Wird der Controller zum Data Scientist? Herausforderungen und Chancen in Zeiten von Big Data, Predictive Analytics und Echtzeitverfügbarkeit. In: B. Hermeier, T. Heupel und S. Fichtner-Rosada (Hg.): *Arbeitswelten der Zukunft. Wie die Digitalisierung unsere Arbeitsplätze und Arbeitsweisen verändert*. Springer Berlin Heidelberg, S. 201–221.

Highhouse, S. (2008): Stubborn Reliance on Intuition and Subjectivity in Employee Selection. In: *Industrial and Organizational Psychology* 1 (03), S. 333–342. DOI: 10.1111/j.1754-9434.2008.00058.x.

Hirst, M. K.; Lowy, S. M. (1990): The linear additive and interactive effects of budgetary goal difficulty and feedback on performance. In: *Accounting, Organizations and Society* 15 (5), S. 425–436. DOI: 10.1016/0361-3682(90)90026-Q.

Hirst, M. K.; Yetton, P. W. (1999): The effects of budget goals and task interdependence on the level of and variance in performance: a research note. In: *Accounting, Organizations and Society* 24 (3), S. 205–216. DOI: 10.1016/S0361-3682(98)00037-3.

Hoff, K. A.; Bashir, M. (2015): Trust in automation: integrating empirical evidence on factors that influence trust. In: *Human Factors* 57 (3), S. 407–434. DOI: 10.1177/0018720814547570.

Holderness, D. K.; Huffman, A.; Lewis-Western, M. (2019): Rank and file equity compensation and earnings management: Evidence from stock options. In: *Journal of Business Finance & Accounting* 46 (9-10), S. 1201–1236. DOI: 10.1111/jbfa.12404.

Holderness, D. K.; Olsen, K. J.; Thornock, T. A. (2017): Who Are You to Tell Me That?! The Moderating Effect of Performance Feedback Source and Psychological Entitlement on Individual Performance. In: *Journal of Management Accounting Research* 29 (2), S. 33–46. DOI: 10.2308/jmar-51538.

Holderness, D. K.; Olsen, K. J.; Thornock, T. A. (2020): Assigned versus Chosen Relative Performance Information: The Effect of Feedback Frequency on Performance. In: *Journal of Management Accounting Research* 32 (1), S. 137–158. DOI: 10.2308/jmar-52504.

Hollenbeck, J. R.; Williams, C. R. (1987): Goal importance, self-focus, and the goal-setting process. In: *Journal of Applied Psychology* 72 (2), S. 204–211. DOI: 10.1037/0021-9010.72.2.204.

Holmbeck, G. N. (1997): Toward terminological, conceptual, and statistical clarity in the study of mediators and moderators: examples from the child-clinical and pediatric psychology literatures. In: *Journal of Consulting and Clinical Psychology* 65 (4), S. 599.

Holthoff, G.; Decher, R. (2020): Implementierung von Predictive Analytics im Forecasting. In: *Controlling* 32 (6), S. 53–59. DOI: 10.15358/0935-0381-2020-6-53.

Holzhacker, M. (2019): Die Schattenseiten innerbetrieblicher Leistungsvergleiche. In: *Controlling & Management Review* 63 (6), S. 50–55. DOI: 10.1007/s12176-019-0040-7.

Horton, J. J.; Rand, D. G.; Zeckhauser, R. J. (2011): The online laboratory: conducting experiments in a real labor market. In: *Experimental Economics* 14 (3), S. 399–425. DOI: 10.1007/s10683-011-9273-9.

Horváth, P. (1978): Controlling - Entwicklung und Stand einer Konzeption zur Lösung der Adaptions- und Koordinationsprobleme der Führung. In: *Journal of Business Economics* 48 (194-208).

Horváth & Partners (2017): Mit Digital Forecasting den Vertrieb besser steuern. Methodik und Nutzen automatisierter Forecasts. Hg. v. Horváth & Partners.

- Hribar, P.; Yang, H. (2016): CEO Overconfidence and Management Forecasting. In: *Contemporary Accounting Research* 33 (1), S. 204–227. DOI: 10.1111/1911-3846.12144.
- Hubbard, R.; Armstrong, J. (1994): Replications and extensions in marketing: Rarely published but quite contrary. In: *International Journal of Research in Marketing* 11 (3), S. 233–248. DOI: 10.1016/0167-8116(94)90003-5.
- Huber, O.; Kunz, U. (2007): Time pressure in risky decision-making: effect on risk defusing. In: *Psychology Science* 49 (4), S. 415.
- Huerta, E.; Jensen, S. (2017): An Accounting Information Systems Perspective on Data Analytics and Big Data. In: *Journal of Information Systems* 31 (3), S. 101–114. DOI: 10.2308/isis-51799.
- Hutchinson, J. W.; Alba, J. W.; Eisenstein, E. M. (2010): Heuristics and Biases in Data-Based Decision Making: Effects of Experience, Training, and Graphical Data Displays. In: *Journal of Marketing Research* 47 (4), S. 627–642. DOI: 10.1509/jmkr.47.4.627.
- Iacobucci, D.; Posavac, S. S.; Kardes, F. R.; Schneider, M. J.; Popovich, D. L. (2015): Toward a more nuanced understanding of the statistical properties of a median split. In: *Journal of Consumer Psychology* 25 (4), S. 652–665. DOI: 10.1016/j.jcps.2014.12.002.
- Ilgen, D.; Davis, C. (2000): Bearing Bad News: Reactions to Negative Performance Feedback. In: *Applied Psychology* 49 (3), S. 550–565. DOI: 10.1111/1464-0597.00031.
- Indjejikian, R. J.; Matějka, M.; Merchant, K. A.; van der Stede, W. A. (2014): Earnings Targets and Annual Bonus Incentives. In: *The Accounting Review* 89 (4), S. 1227–1258. DOI: 10.2308/accr-50732.
- Ireland, L. (2020): Who errs? Algorithm aversion, the source of judicial error, and public support for self-help behaviors. In: *Journal of Crime and Justice* 43 (2), S. 174–192. DOI: 10.1080/0735648X.2019.1655781.
- Jacobson, J.; Dobbs-Marsh, J.; Liberman, V.; Minson, J. A. (2011): Predicting Civil Jury Verdicts: How Attorneys Use (and Misuse) a Second Opinion. In: *Journal of Empirical Legal Studies* 8, S. 99–119. DOI: 10.1111/j.1740-1461.2011.01229.x.

Jung, M.; Beuckes, T.; Ostrowicz, S. (2018): Enterprise Automation Concept - Integrierter Lösungsansatz für robotergesteuerte Prozessautomatisierung. Hg. v. IPRI gGmbH und Horváth & Partners GmbH. Stuttgart.

Jung, M.; Seiter, M. (2021): Towards a better understanding on mitigating algorithm aversion in forecasting: an experimental study. In: *Journal of Management Control* 32 (4), S. 495–516. DOI: 10.1007/s00187-021-00326-3.

Jussupow, E.; Benbasat, I.; and Heinzl, A. (2020): Why are we Averse Towards Algorithms? A Comprehensive Literature Review on Algorithm Aversion. In: *Twenty-Eighth European Conference on Information Systems (ECIS2020)*.

Kappes, M.; Leyk, J. (2018): Digitale Planung. Überblick über die Planung der Zukunft im Zuge der Digitalisierung. In: *Controlling* (06), S. 4–12.

Keding, C.; Meissner, P. (2021): Managerial overreliance on AI-augmented decision-making processes: How the use of AI-based advisory systems shapes choice behavior in R&D investment decisions. In: *Technological Forecasting and Social Change* 171, S. 120970. DOI: 10.1016/j.techfore.2021.120970.

Kelly, T.; Margheim, L.; Pattison, D. (2011): Survey On The Differential Effects Of Time Deadline Pressure Versus Time Budget Pressure On Auditor Behavior. In: *Journal of Applied Business Research* 15 (4), S. 117. DOI: 10.19030/jabr.v15i4.5666.

Kerstholt, J. (1994): The effect of time pressure on decision-making behaviour in a dynamic task environment. In: *Acta Psychologica* 86 (1), S. 89–104.

Kirchler, M.; Andersson, D.; Bonn, C.; Johannesson, M.; Sørensen, E. Ø.; Stefan, M. et al. (2017): The effect of fast and slow decisions on risk taking. In: *Journal of Risk and Uncertainty* 54 (1), S. 37–59. DOI: 10.1007/s11166-017-9252-4.

Klassen, R. D.; Flores, B. E. (2001): Forecasting practices of Canadian firms: Survey results and comparisons. In: *International Journal of Production Economics* 70 (2), S. 163–174. DOI: 10.1016/S0925-5273(00)00063-3.

Klein, H. J.; Whitener, E. M.; Ilgen, D. R. (1990): The role of goal specificity in the goal-setting process. In: *Motivation and Emotion* 14 (3), S. 179–193. DOI: 10.1007/BF00995568.

Kleinberg, J.; Ludwig, J.; Mullainathan, S.; Rambachan, A. (2018): Algorithmic Fairness. In: *AEA Papers and Proceedings* 108, S. 22–27. DOI: 10.1257/pandp.20181018.

Kluger, A. N.; DeNisi, A. (1996): The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. In: *Psychological Bulletin* 119 (2), S. 254–284. DOI: 10.1037/0033-2909.119.2.254.

Kluger, A. N.; DeNisi, A. (1998): Feedback Interventions. In: *Current Directions in Psychological Science* 7 (3), S. 67–72. DOI: 10.1111/1467-8721.ep10772989.

Knauer, T.; Sommer, F.; Wöhrmann, A. (2017): Tournament Winner Proportion and its Effect on Effort: An Investigation of the Underlying Psychological Mechanisms. In: *European Accounting Review* 26 (4), S. 681–702. DOI: 10.1080/09638180.2016.1175957.

Knuth, D. E. (1990): *Fundamental algorithms*. 2. ed., [Nachdr.]. Reading, Mass.: Addison-Wesley (World student series, / Donald E. Knuth ; Vol. 1).

Kocher, M. G.; Schindler, D.; Trautmann, S. T.; Xu, Y. (2019): Risk, time pressure, and selection effects. In: *Experimental Economics* 22 (1), S. 216–246. DOI: 10.1007/s10683-018-9576-1.

Kocher, M. G.; Sutter, M. (2006): Time is money—Time pressure, incentives, and the quality of decision-making. In: *Journal of Economic Behavior & Organization* 61 (3), S. 375–392. DOI: 10.1016/j.jebo.2004.11.013.

Kokina, J.; Blanchette, S. (2019): Early evidence of digital labor in accounting: Innovation with Robotic Process Automation. In: *International Journal of Accounting Information Systems* 35, S. 100431. DOI: 10.1016/j.accinf.2019.100431.

Köszegi, B. (2006): Ego utility, overconfidence, and task choice. In: *Journal of the European Economic Association* 4 (4), S. 673–707. DOI: 10.1162/JEEA.2006.4.4.673.

Kramer, S.; Maas, V. S.; van Rinsum, M. (2016): Relative performance information, rank ordering and employee performance: A research note. In: *Management Accounting Research* 33, S. 16–24. DOI: 10.1016/j.mar.2016.03.004.

Kromrey, H.; Roose, J.; Strübing, J. (2016): *Empirische Sozialforschung. Modelle und Methoden der standardisierten Datenerhebung und Datenauswertung mit Annotationen aus qualitativ-interpretativer Perspektive*. 13., völlig überarbeitete Auflage. Konstanz: UVK Verlagsgesellschaft mbH.

Kuckartz, U.; Rädiker, S.; Ebert, T.; Schehl, J. (2010): Kreuztabelle, Chi-Quadrat und Zusammenhangsmaße. In: Udo Kuckartz, Stefan Rädiker, Thomas Ebert und Julia

Schehl (Hg.): Statistik. Eine verständliche Einführung. 1. Aufl. Wiesbaden: VS Verl. für Sozialwiss, S. 81–102.

Kuhnen, C. M.; Tymula, A. (2012): Feedback, Self-Esteem, and Performance in Organizations. In: *Management Science* 58 (1), S. 94–113. DOI: 10.1287/mnsc.1110.1379.

Kuncel, N. R.; Klieger, D. M.; Connelly, B. S.; Ones, D. S. (2013): Mechanical versus clinical data combination in selection and admissions decisions: a meta-analysis. In: *The Journal of applied psychology* 98 (6), S. 1060–1072. DOI: 10.1037/a0034156.

Lambert, T. A.; Jones, K. L.; Brazel, J. F.; Showalter, D. S. (2017): Audit time pressure and earnings quality: An examination of accelerated filings. In: *Accounting, Organizations and Society* 58, S. 50–66. DOI: 10.1016/j.aos.2017.03.003.

Latham, G. P.; Locke, E. A. (1991): Self-regulation through goal setting. In: *Organizational Behavior and Human Decision Processes* 50 (2), S. 212–247. DOI: 10.1016/0749-5978(91)90021-K.

Latham, G. P.; Locke, E. A. (2013): Potential pitfalls in goal setting and how to avoid them. In: *New developments in goal setting and task performance*. New York, NY, US: Routledge/Taylor & Francis Group, S. 569–579.

Lawrence, M.; Goodwin, P.; O'Connor, M.; Önkal, D. (2006): Judgmental forecasting: A review of progress over the last 25 years. In: *International Journal of Forecasting* 22 (3), S. 493–518. DOI: 10.1016/j.ijforecast.2006.03.007.

Lee, J. D.; See, K. A. (2004): Trust in automation: designing for appropriate reliance. In: *Human Factors* 46 (1), S. 50–80. DOI: 10.1518/hfes.46.1.50\_30392.

Legris, P.; Ingham, J.; Colletette, P. (2003): Why do people use information technology? A critical review of the technology acceptance model. In: *Information & Management* 40 (3), S. 191–204. DOI: 10.1016/S0378-7206(01)00143-4.

Leone, A. J.; Rock, S. (2002): Empirical tests of budget ratcheting and its effect on managers' discretionary accrual choices. In: *Journal of Accounting and Economics* 33 (1), S. 43–67. DOI: 10.1016/S0165-4101(01)00044-1.

Libby, R.; Luft, J. (1993): Determinants of judgment performance in accounting settings: Ability, knowledge, motivation, and environment. In: *Accounting, Organizations and Society* 18 (5), S. 425–450. DOI: 10.1016/0361-3682(93)90040-D.

- Liberman, V.; Minson, J. A.; Bryan, C. J.; Ross, L. (2012): Naïve realism and capturing the “wisdom of dyads”. In: *Journal of Experimental Social Psychology* 48 (2), S. 507–512. DOI: 10.1016/j.jesp.2011.10.016.
- Liu, P.; Li, Z. (2012): Task complexity: A review and conceptualization framework. In: *International Journal of Industrial Ergonomics* 42 (6), S. 553–568. DOI: 10.1016/j.ergon.2012.09.001.
- Locke, E. A.; Latham, G. P. (1990): A theory of goal setting & task performance. Englewood Cliffs, N.J.: Prentice Hall.
- Locke, E. A.; Latham, G. P. (2002): Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. In: *American Psychologist* 57 (9), S. 705–717. DOI: 10.1037//0003-066X.57.9.705.
- Locke, E. A.; Latham, G. P. (2006): New Directions in Goal-Setting Theory. In: *Current Directions in Psychological Science* 15 (5), S. 265–268. DOI: 10.1111/j.1467-8721.2006.00449.x.
- Locke, E. A.; Latham, G. P. (2013): Goal Setting Theory, 1990. In: Edwin A. Locke und Gary P. Latham (Hg.): *New developments in goal setting and task performance*. New York, NY: Routledge, S. 3–16.
- Locke, E. A.; Shaw, K. N.; SAARI, L. M.; Latham, G. P. (1981): Goal setting and task performance: 1969-1980. In: *Psychological Bulletin* 90 (1), S. 125–152. DOI: 10.1037/0033-2909.90.1.125.
- Logg, J. M.; Haran, U.; Moore, D. A. (2018): Is overconfidence a motivated bias? Experimental evidence. In: *Journal of Experimental Psychology: General* 147 (10), S. 1445–1465. DOI: 10.1037/xge0000500.
- Logg, J. M.; Minson, J. A.; Moore, D. A. (2019): Algorithm appreciation. People prefer algorithmic to human judgment. In: *Organizational Behavior and Human Decision Processes* 151, S. 90–103. DOI: 10.1016/j.obhdp.2018.12.005.
- Lourenco, S. (2019): Field Experiments in Managerial Accounting Research. In: *Foundations and Trends in Accounting* 14 (1), S. 1–72. DOI: 10.1561/14000000059.
- Lourenco, S.; Greenberg, J. O.; Littlefield, M.; Bates, D. W.; Narayanan, V. G. (2018): The performance effect of feedback in a context of negative incentives: Evidence from

- a field experiment. In: *Management Accounting Research* 40, S. 1–14. DOI: 10.1016/j.mar.2018.01.002.
- Ludwig, K. (2020): Die Politik hat die Virusangst unterschätzt. In: *Süddeutsche Zeitung*, 06.03.2020. Online verfügbar unter <https://www.sueddeutsche.de/politik/corona-virus-krise-hamsterkaeufer-1.4832408>, zuletzt geprüft am 31.05.2022.
- Luft, J. (2016): Management accounting in the laboratory and in social context: Four contrasts, 1975–2014. In: *Accounting, Organizations and Society* 49, S. 9–20. DOI: 10.1016/j.aos.2015.08.001.
- Luhmann, N. (2000): Familiarity, Confidence, Trust: Problems and Alternatives. In: Di-ego Gambetta (Hg.): *Trust: Making and Breaking Cooperative Relations*. electronic edition, S. 94–107.
- MacKinnon, D. P.; Fairchild, A. J.; Fritz, M. S. (2007): Mediation analysis. In: *Annual Review of Psychology* 58, S. 593–614. DOI: 10.1146/an-nurev.psych.58.110405.085542.
- MacKinnon, D. P.; Lockwood, C. M.; Hoffman, J. M.; West, S. G.; Sheets, V. (2002): A comparison of methods to test mediation and other intervening variable effects. In: *Psychological Methods* 7 (1), S. 83–104. DOI: 10.1037/1082-989x.7.1.83.
- Magretta, J. (2002): The Behavior behind the Buzzwords. In: *MIT Sloan Management Review* 43 (4), S. 89–93.
- Mahlendorf, M. D.; Kleinschmit, F.; Perego, P. (2014): Relational effects of relative performance information: The role of professional identity. In: *Accounting, Organizations and Society* 39 (5), S. 331–347. DOI: 10.1016/j.aos.2014.05.001.
- Makridakis, S.; Andersen, A.; Carbone, R.; Fildes, R.; Hibon, M.; Lewandowski, R. et al. (1982): The accuracy of extrapolation (time series) methods: Results of a forecast-ing competition. In: *Journal of Forecasting* 1 (2), S. 111–153. DOI: 10.1002/for.3980010202.
- Malmi, T.; Brown, D. A. (2008): Management control systems as a package—Oppor-tunities, challenges and research directions. In: *Management Accounting Research* 19 (4), S. 287–300. DOI: 10.1016/j.mar.2008.09.003.

Marangunić, N.; Granić, A. (2015): Technology acceptance model: a literature review from 1986 to 2013. In: *Universal Access in the Information Society* 14 (1), S. 81–95. DOI: 10.1007/s10209-014-0348-1.

Marginson, D.; Ogden, S. (2005): Coping with ambiguity through the budget: the positive effects of budgetary targets on managers' budgeting behaviours. In: *Accounting, Organizations and Society* 30 (5), S. 435–456. DOI: 10.1016/j.aos.2004.05.004.

Mason, W.; Suri, S. (2012): Conducting behavioral research on Amazon's Mechanical Turk. In: *Behavior Research Methods* 44 (1), S. 1–23. DOI: 10.3758/s13428-011-0124-6.

Matanovic, S.; Sabel, R.; Wöhrmann, A. (2019): Controller vs. Data Scientist: Eine empirische Analyse des Status quo der Aufgaben. In: *Controlling* 31 (6), S. 54–62. DOI: 10.15358/0935-0381-2019-6-54.

Matt, C.; Hess, T.; Benlian, A.; Wiesbock, F. (2016): Options for Formulating a Digital Transformation Strategy. In: *MIS Quarterly Executive* 15 (2), S. 103–119.

Maule, A.; Hockey, G. J.; Bdzola, L. (2000): Effects of time-pressure on decision-making under uncertainty: changes in affective state and information processing strategy. In: *Acta Psychologica* 104 (3), S. 283–301. DOI: 10.1016/S0001-6918(00)00033-0.

McAllister, D. J. (1995): Affect- and Cognition-Based Trust as Foundations for Interpersonal Cooperation in Organizations. In: *Academy of Management Journal* 38 (1), S. 24–59. DOI: 10.5465/256727.

McCarty, P. A. (1986): Effects of Feedback on the Self-Confidence of Men and Women. In: *Academy of Management Journal* 29 (4), S. 840–847. DOI: 10.2307/255950.

McKinney, E.; Yoos, C. J.; Snead, K. (2017): The need for 'skeptical' accountants in the era of Big Data. In: *Journal of Accounting Education* 38, S. 63–80. DOI: 10.1016/j.jaccedu.2016.12.007.

McKinsey (2017): Smartening up with Artificial Intelligence (AI) - What's in it for Germany and its Industrial Sector? Hg. v. McKinsey. Online verfügbar unter <https://www.mckinsey.com/~media/mckinsey/industries/semiconductors/our%20insights/smartening%20up%20with%20artificial%20intelligence/smartening-up-with-artificial-intelligence.ashx>, zuletzt geprüft am 31.05.2022.

- Meehl, P. E. (1954): Clinical versus statistical prediction. A theoretical analysis and a review of the evidence. University of Minnesota Press: Echo Point Books & Media.
- Mehanna, W.; Tatzel, J.; Vogel, P. (2018): Business Analytics im Controlling - Fünf Anwendungsfehler. In: *Controlling* 30, S. 502–508. DOI: 10.15358/0935-0381-2018-S-38.
- Merchant, K. A.; Otley, D. T. (2006): A Review of the Literature on Control and Accountability. In: *Handbook of Management Accounting Research* 2, S. 785–802. DOI: 10.1016/S1751-3243(06)02013-X.
- Merchant, K. A.; van der Stede, W. A. (2017): Management control systems. Performance measurement, evaluation, and incentives. Fourth edition. Harlow, England: Pearson.
- Meub, L.; Proeger, T. (2016): Can anchoring explain biased forecasts? Experimental evidence. In: *Journal of Behavioral and Experimental Finance* 12, S. 1–13. DOI: 10.1016/j.jbef.2016.08.001.
- Mischer, M.; Kirsch, T.; Günther, T. (2021): Die ganzheitliche Bewertung von Zeitmanagementmaßnahmen durch Verknüpfung „weicher“ und „harter“ Kennzahlen. In: *Controlling* 33 (6), S. 19–28.
- Mitra, A.; Gupta, N.; Jenkins, G. D., JR. (1997): A drop in the bucket: when is a pay raise a pay raise? In: *Journal of Organizational Behavior* 18 (2), S. 117–137. DOI: 10.1002/(SICI)1099-1379(199703)18:2<117::AID-JOB790>3.0.CO;2-1.
- Möller, K.; Schäffer, U.; Verbeeten, F. (2020): Digitalization in management accounting and control: an editorial. In: *Journal of Management Control* 31 (1-2), S. 1–8. DOI: 10.1007/s00187-020-00300-5.
- Moore, D. A.; Healy, P. J. (2008): The trouble with overconfidence. In: *Psychological Review* 115 (2), S. 502–517. DOI: 10.1037/0033-295X.115.2.502.
- Morse, S.; Gergen, K. J. (1970): Social comparison, self-consistency, and the concept of self. In: *Journal of Personality and Social Psychology* 16 (1), S. 148–156. DOI: 10.1037/h0029862.
- Mussweiler, T.; Rüter, K.; Epstude, K. (2004): The ups and downs of social comparison: mechanisms of assimilation and contrast. In: *Journal of Personality and Social Psychology* 87 (6), S. 832–844. DOI: 10.1037/0022-3514.87.6.832.

- Nagar, V. (2002): Delegation and Incentive Compensation. In: *The Accounting Review* 77 (2), S. 379–395. DOI: 10.2308/accr.2002.77.2.379.
- Nandhakumar, J.; Jones, M. (2001): Accounting for time: managing time in project-based teamworking. In: *Accounting, Organizations and Society* 26 (3), S. 193–214. DOI: 10.1016/S0361-3682(99)00051-3.
- Newman, A. H.; Stikeleather, B. R.; Waddoups, N. J. (2021): How Relative Performance Information Affects Exploration-exploitation Decisions. In: *Journal of Management Accounting Research*, S. 75–95. DOI: 10.2308/JMAR-2020-070.
- Niederer, A. (2020): Das neue Coronavirus ist ansteckender als zunächst gedacht – aber es ist bekämpfbar. In: *Neue Züricher Zeitung*, 21.01.2020. Online verfügbar unter <https://www.nzz.ch/meinung/das-neue-coronavirus-ist-ansteckender-als-zunaechst-gedacht-aber-es-ist-bekaempfbar-ld.1535365>, zuletzt geprüft am 31.05.2022.
- Nihalani, P. K.; Mayrath, M.; Robinson, D. H. (2011): When feedback harms and collaboration helps in computer simulation environments: An expertise reversal effect. In: *Journal of Educational Psychology* 103 (4), S. 776–785. DOI: 10.1037/a0025276.
- Niszczota, P.; Kaszás, D. (2020): Robo-investment aversion. In: *PloS one* 15 (9), e0239277. DOI: 10.1371/journal.pone.0239277.
- Nixon, W. A.; Burns, J. (2005): Management control in the 21st century. In: *Management Accounting Research* 16 (3), S. 260–268. DOI: 10.1016/j.mar.2005.07.001.
- Obermaier, R.; Müller, F. (2008): Management accounting research in the lab – method and applications. In: *Zeitschrift für Planung & Unternehmenssteuerung* 19 (3), S. 325–351. DOI: 10.1007/s00187-008-0056-1.
- O'Conner, S. (2021): Why ranking employees by performance backfires. Online verfügbar unter <https://www.ft.com/content/0691002c-2200-4583-88c9-9c942d534228>, zuletzt geprüft am 31.05.2022.
- Ordóñez, L. D.; Schweitzer, M. E.; Galinsky, A. D.; Bazerman, M. H. (2009): Goals Gone Wild: The Systematic Side Effects of Overprescribing Goal Setting. In: *Academy of Management Perspectives* 23 (1), S. 6–16. DOI: 10.5465/amp.2009.37007999.
- O'reilly, J.; Aquino, K. (2011): A Model of Third Parties' Morally Motivated Responses to Mistreatment in Organizations. In: *Academy of Management Review* 36 (3), S. 526–543. DOI: 10.5465/amr.2009.0311.

- Oshri, I.; Plugge, A.: What Do You See in Your Bot? Lessons from KAS Bank. In: *Digital Technologies for Global Sourcing of Services* 410, S. 145–161. DOI: 10.1007/978-3-030-66834-1\_9.
- Paolacci, G.; Chandler, J.; & Ipeirotis, P. G. (2010): Running experiments on Amazon Mechanical Turk. In: *Judgment and Decision Making* 5 (5), S. 411–419.
- Peel, M. J. (2018): Addressing Unobserved Selection Bias in Accounting Studies: The Bias Minimization Method. In: *European Accounting Review* 27 (1), S. 173–183. DOI: 10.1080/09638180.2016.1220322.
- Pepitone, J. (2013): Microsoft kills employee-ranking system. Online verfügbar unter <https://money.cnn.com/2013/11/13/technology/enterprise/microsoft-stack-ranking/index.html>, zuletzt geprüft am 31.05.2022.
- Petropoulos, F.; Fildes, R.; Goodwin, P. (2016): Do 'big losses' in judgmental adjustments to statistical forecasts affect experts' behaviour? In: *European Journal of Operational Research* 249 (3), S. 842–852. DOI: 10.1016/j.ejor.2015.06.002.
- Pietsch, C. P. R.; Messier, W. F. (2017): The Effects of Time Pressure on Belief Revision in Accounting: A Review of Relevant Literature within a Pressure-Arousal-Effort-Performance Framework. In: *Behavioral Research in Accounting* 29 (2), S. 51–71. DOI: 10.2308/bria-51756.
- Podsakoff, P. M.; Farh, J.-L. (1989): Effects of feedback sign and credibility on goal setting and task performance. In: *Organizational Behavior and Human Decision Processes* 44 (1), S. 45–67. DOI: 10.1016/0749-5978(89)90034-4.
- Polzer, J. T.; Neale, M. A. (1995): Constraints or Catalysts? Reexamining goal Setting Within the Context of Negotiation. In: *Human Performance* 8 (1), S. 3–26. DOI: 10.1207/s15327043hup0801\_2.
- Prahl, A.; van Swol, L. (2017): Understanding algorithm aversion: When is advice from automation discounted? In: *Journal of Forecasting* 36 (6), S. 691–702. DOI: 10.1002/for.2464.
- Preacher, K. J.; Hayes, A. F. (2004): SPSS and SAS procedures for estimating indirect effects in simple mediation models. In: *Behavior Research Methods, Instruments, & Computers* 36 (4), S. 717–731.

Preacher, K. J.; Rucker, D. D.; Hayes, A. F. (2007): Addressing Moderated Mediation Hypotheses: Theory, Methods, and Prescriptions. In: *Multivariate Behavioral Research* 42 (1), S. 185–227. DOI: 10.1080/00273170701341316.

PwC Europe (2018): Auswirkungen der Nutzung von künstlicher Intelligenz in Deutschland. Executive Summary und Empfehlungen. Hg. v. PwC Europe.

Pyszczynski, T.; Greenberg, J.; LaPrelle, J. (1985): Social comparison after success and failure: Biased search for information consistent with a self-serving conclusion. In: *Journal of Experimental Social Psychology* 21 (2), S. 195–211. DOI: 10.1016/0022-1031(85)90015-0.

Quattrone, P. (2016): Management accounting goes digital: Will the move make it wiser? In: *Management Accounting Research* 31, S. 118–122. DOI: 10.1016/j.mar.2016.01.003.

Rasch, D.; Kubinger, K. D.; Moder, K. (2011): The two-sample t test: pre-testing its assumptions does not pay off. In: *Statistical Papers* 52 (1), S. 219–231. DOI: 10.1007/s00362-009-0224-x.

Remus, W.; O'Connor, M.; Griggs, K. (1995): Does reliable information improve the accuracy of judgmental forecasts? In: *International Journal of Forecasting* 11 (2), S. 285–293. DOI: 10.1016/0169-2070(94)00578-Z.

Renier, L. A.; Schmid Mast, M.; Bekbergenova, A. (2021): To err is human, not algorithmic – Robust reactions to erring algorithms. In: *Computers in Human Behavior* 124, S. 106879. DOI: 10.1016/j.chb.2021.106879.

Rikhardsson, P.; Yigitbasioglu, O. (2018): Business intelligence & analytics in management accounting research. Status and future focus. In: *International Journal of Accounting Information Systems* 29, S. 37–58. DOI: 10.1016/j.accinf.2018.03.001.

Rogers, H. (1987): Theory of recursive functions and effective computability. 5. print. Cambridge, Mass.: MIT Press.

Rose, D.; Justine, L. (2013): Current Practices in 360 Degree Feedback: A Benchmark Study of North American Companies. 3D Group. Emeryville.

Ross, J.; Irani, L.; Silberman, M. S.; Zaldivar, A.; Tomlinson, B. (2010): Who are the crowdworkers? In: *Extended Abstracts on Human Factors in Computing Systems*, S. 2863–2872. DOI: 10.1145/1753846.1753873.

- Rousseau, D. M.; Sitkin, S. B.; Burt, R. S.; Camerer, C. (1998): Not so different after all: A cross-discipline view of trust. In: *The Academy of Management Review* 23 (3), S. 393–404.
- Sarris, V. (1990): Erkenntnisgewinnung und Methodik der experimentellen Psychologie. München: UTB / BRO.
- Schäffer, U.; Weber, J. (2016): Die Digitalisierung wird das Controlling radikal verändern. In: *Controlling & Management Review* (6), S. 8–17.
- Schläfke, M.; Silvi, R.; Möller, K. (2013): A framework for business analytics in performance management. In: *International Journal of Productivity and Performance Management* 62 (1), S. 110–122. DOI: 10.1108/17410401311285327.
- Seckler, C.; Gronewold, U.; Reihlen, M. (2017): An error management perspective on audit quality: Toward a multi-level model. In: *Accounting, Organizations and Society* 62 (1), S. 21–42. DOI: 10.1016/j.aos.2017.08.004.
- Seijts, G. H.; Latham, G. P.; Woodwark, M. (2013): Learning Goals: A Qualitative and Quantitative Review. In: Edwin A. Locke und Gary P. Latham (Hg.): *New developments in goal setting and task performance*. New York, NY: Routledge, S. 195–212.
- Seiter, M. (2019): *Business Analytics. Wie Sie Daten für die Steuerung von Unternehmen nutzen*. 2., komplett überarbeitete und erweiterte Auflage. München: Verlag Franz Vahlen.
- Sejdić, G. (2020): Digitale Transformation des Controlling/Finance-Bereichs. In: *Controlling* 32 (4), S. 61–67. DOI: 10.15358/0935-0381-2020-4-61.
- Shariff, A.; Bonnefon, J.-F.; Rahwan, I. (2021): How safe is safe enough? Psychological mechanisms underlying extreme safety demands for self-driving cars. In: *Transportation Research Part C: Emerging Technologies* 126, S. 103069. DOI: 10.1016/j.trc.2021.103069.
- Shaw, J. D.; Gupta, N. (2015): Let the evidence speak again! Financial incentives are more effective than we thought. In: *Human Resource Management Journal* 25 (3), S. 281–293. DOI: 10.1111/1748-8583.12080.
- Sibbertsen, P.; Lehne, H. (2021): *Statistik. Einführung für Wirtschafts- und Sozialwissenschaftler*. 3., korrigierte und überarbeitete Auflage. Berlin, Heidelberg: Springer Gabler.

Silverman, R. (2011): Latest game theory: Mixing work and play. In: *The Wall Street Journal*, 2011. Online verfügbar unter <https://www.wsj.com/articles/SB10001424052970204294504576615371783795248>, zuletzt geprüft am 31.05.2022.

Smith, A. (1759): *The Theory of Moral Sentiments*.

Smyth, J.; Pearson, J. (2011): Internet Survey Methods: A Review of Strengths, Weakness, and Innovations. In: Marcel Das, Peter Ester und Lars Kaczmirek (Hg.): *Social and behavioral research and the internet. Advances in applied methods and research strategies*. New York, NY: Routledge (European Association of Methodology series), S. 11–45.

Snizek, J. A.; Buckley, T. (1995): Cueing and Cognitive Conflict in Judge-Advisor Decision Making. In: *Organizational Behavior and Human Decision Processes* 62 (2), S. 159–174. DOI: 10.1006/obhd.1995.1040.

Soll, J. B.; Larrick, R. P. (2009): Strategies for revising judgment: how (and how well) people use others' opinions. In: *Journal of Experimental Psychology. Learning, Memory, and Cognition* 35 (3), S. 780–805. DOI: 10.1037/a0015145.

Song, H.; Tucker, A. L.; Murrell, K. L.; Vinson, D. R. (2018): Closing the Productivity Gap: Improving Worker Productivity Through Public Relative Performance Feedback and Validation of Best Practices. In: *Management Science* 64 (6), S. 2628–2649. DOI: 10.1287/mnsc.2017.2745.

Speckbacher, G.; Wabnegg, M. (2020): Incentivizing innovation: The role of knowledge exchange and distal search behavior. In: *Accounting, Organizations and Society* 86, S. 101142. DOI: 10.1016/j.aos.2020.101142.

Spiliopoulos, L.; Ortmann, A. (2018): The BCD of response time analysis in experimental economics. In: *Experimental Economics* 21 (2), S. 383–433. DOI: 10.1007/s10683-017-9528-1.

Spilker, B. (1995): The Effects of Time Pressure and Knowledge on Key Word Selection Behavior in Tax Research. In: *Accounting Review* 70 (1), S. 49–70.

- Stewart, G. L.; Courtright, S. H.; Barrick, M. R. (2012): Peer-based control in self-managing teams: linking rational and normative influence with individual and group performance. In: *The Journal of applied psychology* 97 (2), S. 435–447. DOI: 10.1037/a0025303.
- Strauß, E.; Zecher, C. (2013): Management control systems: a review. In: *Journal of Management Control* 23 (4), S. 233–268. DOI: 10.1007/s00187-012-0158-7.
- Suls, J.; Wheeler, L. (2000): Handbook of Social Comparison. Boston, MA: Springer US.
- Sun, Y.; Xu, W. (2012): The role of accounting conservatism in management forecast bias. In: *Journal of Contemporary Accounting & Economics* 8 (2), S. 64–77. DOI: 10.1016/j.jcae.2012.05.002.
- Svanström, T. (2016): Time Pressure, Training Activities and Dysfunctional Auditor Behaviour: Evidence from Small Audit Firms. In: *International Journal of Auditing* 20 (1), S. 42–51. DOI: 10.1111/ijau.12054.
- Tafkov, I. D. (2013): Private and Public Relative Performance Information under Different Compensation Contracts. In: *The Accounting Review* 88 (1), S. 327–350. DOI: 10.2308/accr-50292.
- Tagesschau (2020): RKI warnt vor Millionen Infektionen. Tagesschau. Online verfügbar unter <https://www.tagesschau.de/inland/corona-rki-101.html>, zuletzt geprüft am 31.05.2022.
- Teoh, S. H. (2018): The promise and challenges of new datasets for accounting research. In: *Accounting, Organizations and Society* 68-69, S. 109–117. DOI: 10.1016/j.aos.2018.03.008.
- Thornock, T. A. (2016): How the timing of performance feedback impacts individual performance. In: *Accounting, Organizations and Society* 55, S. 1–11. DOI: 10.1016/j.aos.2016.09.002.
- Tiron-Tudor, A.; Deliu, D. (2021): Big Data's Disruptive Effect on Job Profiles: Management Accountants' Case Study. In: *Journal of Risk and Financial Management* 14 (8), S. 376. DOI: 10.3390/jrfm14080376.

- Tran, A.; Zeckhauser, R. (2012): Rank as an inherent incentive: Evidence from a field experiment. In: *Journal of Public Economics* 96 (9-10), S. 645–650. DOI: 10.1016/j.jpubeco.2012.05.004.
- Tuttle, B.; Collier, M.; Burton, F. (1997): An examination of market efficiency: Information order effects in a laboratory market. In: *Accounting, Organizations and Society* 22 (1), S. 89–103. DOI: 10.1016/S0361-3682(96)00026-8.
- Tversky, A.; Kahneman, D. (1974): Judgment under Uncertainty: Heuristics and Biases. In: *Science* 185 (4157), S. 1124–1131. DOI: 10.1126/science.185.4157.1124.
- Utley, J. (2014): Reducing Bias in Hierarchical Forecasting. In: *Advances in Business and Management Forecasting* 10, S. 131–141. DOI: 10.1108/S1477-407020140000010019.
- van Dyck, C.; Frese, M.; Baer, M.; Sonnentag, S. (2005): Organizational error management culture and its impact on performance: a two-study replication. In: *Journal of Applied Psychology* 90 (6), S. 1228–1240. DOI: 10.1037/0021-9010.90.6.1228.
- Vancouver, J. B.; Tischner, E. C. (2004): The effect of feedback sign on task performance depends on self-concept discrepancies. In: *Journal of Applied Psychology* 89 (6), S. 1092–1098. DOI: 10.1037/0021-9010.89.6.1092.
- Venkatesh; Morris; Davis (2003): User Acceptance of Information Technology: Toward a Unified View. In: *MIS Quarterly* 27 (3), S. 425. DOI: 10.2307/30036540.
- Venkatesh, V.; Bala, H. (2008): Technology Acceptance Model 3 and a Research Agenda on Interventions. In: *Decision Sciences* 39 (2), S. 273–315. DOI: 10.1111/j.1540-5915.2008.00192.x.
- Venkatesh, V.; Davis, F. D. (2000): A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. In: *Management Science* 46 (2), S. 186–204. DOI: 10.1287/mnsc.46.2.186.11926.
- Verbeeten, F. H. (2008): Performance management practices in public sector organizations. In: *Accounting, Auditing & Accountability Journal* 21 (3), S. 427–454. DOI: 10.1108/09513570810863996.
- Vitale, G.; Cupertino, S.; Riccaboni, A. (2020): Big data and management control systems change: the case of an agricultural SME. In: *Journal of Management Control* 5 (2), S. 107. DOI: 10.1007/s00187-020-00298-w.

- Waldforst, S. (2008): Die Wirkung von Zielen auf die Arbeitsleistung von Akteuren. 1. Aufl. s.l.: Gabler Verlag (Research in Management Accounting & Control).
- Watson, H. J. (2017): Preparing for the Cognitive Generation of Decision Support. In: *MIS Quarterly Executive* 16 (3), S. 153–169.
- Webb, A.; Jeffrey, S. A.; Schulz, A. (2010): Factors Affecting Goal Difficulty and Performance When Employees Select Their Own Performance Goals. Evidence from the Field. In: *Journal of Management Accounting Research* 22 (1), S. 209–232. DOI: 10.2308/jmar.2010.22.1.209.
- Webb, R. A.; Williamson, M. G.; Zhang, Y. ((2013): Productivity-Target Difficulty, Target-Based Pay, and Outside-the-Box Thinking. In: *The Accounting Review* 88 (4), S. 1433–1457. DOI: 10.2308/accr-50436.
- Weber, J.; Schäffer, U. (2001): Controlling als Rationalitätssicherung der Führung — Zum Stand unserer Forschung. In: Jürgen Weber und Utz Schäffer (Hg.): Rationalitätssicherung der Führung. Beiträge zu einer Theorie des Controlling. Wiesbaden: Deutscher Universitätsverlag (Schriften des Center for Controlling & Management (CCM), 2), S. 1–6.
- Wegier, P.; Spaniol, J. (2015): The effect of time pressure on risky financial decisions from description and decisions from experience. In: *PloS one* 10 (4), e0123740. DOI: 10.1371/journal.pone.0123740.
- Weißberger, B. (2021): Künstliche Intelligenz als Zukunftstechnologie im Controlling. In: *Controlling & Management Review* 65, S. 8–17.
- Weitzman, M. (1980): The "Ratchet Principle" and Performance Incentives. In: *Bell Journal of Economics* 11 (1), S. 302–308.
- Wicklund, R. A. (1975): Objective self-awareness. In: *Advances in Experimental Social Psychology* 8, S. 233–275.
- Wildermuth, V.; Blumenthal, U. (2020): Warum selbst vage Prognosen nützlich sein können. Coronavirus-Epidemie. In: *Deutschlandfunk*, 24.02.2020. Online verfügbar unter <https://www.deutschlandfunk.de/coronavirus-epidemie-warum-selbst-vage-prognosen-nuetzlich-100.html>, zuletzt geprüft am 31.05.2022.
- Wills, T. A. (1981): Downward comparison principles in social psychology. In: *Psychological Bulletin* 90 (2), S. 245–271. DOI: 10.1037/0033-2909.90.2.245.

Wirthmann, L. (2014): Is flat better? Zappos ditches hierarchy to improve company performance. In: *Forbes*, 2014.

Woike, J. K.; Hafenbrädl, S. (2020): Rivals without a cause? Relative performance feedback creates destructive competition despite aligned incentives. In: *Journal of Behavioral Decision Making* 33 (4), S. 523–537. DOI: 10.1002/bdm.2162.

Wood, R.; Bandura, A.; Bailey, T. (1990): Mechanisms governing organizational performance in complex decision-making environments. In: *Organizational Behavior and Human Decision Processes* 46 (2), S. 181–201. DOI: 10.1016/0749-5978(90)90028-8.

Yates, D.; Moore, D.; McCabe, G. (1999): *The Practice of Statistics*. New York: Freeman.

Zakay, D.; Wooler, S. (1984): Time pressure, training and decision effectiveness. In: *Ergonomics* 27 (3), S. 273–284. DOI: 10.1080/00140138408963489.

## 8 Anhang

Anhang 1: Identifizierte Artikel zum Stand der Forschung

<b>Autor und Jahr</b>	<b>Titel</b>
<b>Dietvorst et al. (2015)</b>	Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err
<b>Prahl und van Swol (2017)</b>	Understanding algorithm aversion: When is advice from automation discounted?
<b>Dietvorst et al. (2018)</b>	Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them
<b>Logg et al. (2019)</b>	Algorithm appreciation - People prefer algorithmic to human judgment
<b>Castelo et al. (2019)</b>	Task-Dependent Algorithm Aversion
<b>Burton et al. (2019)</b>	A systematic review of algorithm aversion in augmented decision making
<b>Efendić et al. (2020)</b>	Slow response times undermine trust in algorithmic (but not human) predictions
<b>Feng und Gao (2020)</b>	Is optimal recommendation the best? A laboratory investigation under the newsvendor problem
<b>Ireland (2020)</b>	Who errs? Algorithm aversion, the source of judicial error, and public support for self-help behaviors
<b>Keding und Meissner (2021)</b>	Managerial overreliance on AI-augmented decision-making processes: How the use of AI-based advisory systems shapes choice behavior in R&D investment decisions
<b>Jung und Seiter (2021)</b>	Towards a better understanding on mitigating algorithm aversion in forecasting: an experimental study
<b>Filiz et al. (2021)</b>	Reducing algorithm aversion through experience
<b>Berger et al. (2021)</b>	Watch Me Improve-Algorithm Aversion and Demonstrating the Ability to Learn
<b>Renier et al. (2021)</b>	To err is human, not algorithmic-Robust reactions to erring algorithms
<b>Shariff et al. (2021)</b>	How safe is safe enough? Psychological mechanisms underlying extreme safety demands for self-driving cars
<b>Niszczoła und Kaszás (2020)</b>	Robo-investment aversion

## Algorithmen-Aversion im Controlling

### Anhang 2: Forecasts des Algorithmus

<b>Staat</b>	<b>Forecast</b>	<b>Tatsächlicher Rang</b>	<b>Forecastinggenauigkeit</b>
<b>Alabama</b>	34	34	0
<b>Alaska</b>	44	29	15
<b>Arizona</b>	21	9	12
<b>Arkansas</b>	35	37	2
<b>California</b>	1	1	0
<b>Colorado</b>	13	7	6
<b>Connecticut</b>	27	33	6
<b>Delaware</b>	50	50	0
<b>Florida</b>	1	3	2
<b>Georgia</b>	7	4	3
<b>Hawaii</b>	26	17	9
<b>Idaho</b>	44	38	6
<b>Illinois</b>	4	5	1
<b>Indiana</b>	27	28	1
<b>Iowa</b>	27	40	13
<b>Kansas</b>	28	45	17
<b>Kentucky</b>	28	26	2
<b>Louisiana</b>	28	27	1
<b>Maine</b>	49	43	6
<b>Maryland</b>	13	20	7
<b>Massachusetts</b>	15	16	1
<b>Michigan</b>	13	14	1
<b>Minnesota</b>	13	15	2
<b>Mississippi</b>	34	42	8
<b>Missouri</b>	22	19	3
<b>Montana</b>	45	39	6
<b>Nebraska</b>	34	35	1
<b>Nevada</b>	7	11	4
<b>New Hampshire</b>	35	41	6
<b>New Jersey</b>	10	18	8
<b>New Mexico</b>	42	32	10
<b>New York</b>	4	6	2
<b>North Carolina</b>	10	8	2
<b>North Dakota</b>	49	44	5
<b>Ohio</b>	20	23	3
<b>Oklahoma</b>	34	31	3
<b>Oregon</b>	24	24	0
<b>Pennsylvania</b>	7	12	5
<b>Rhode Island</b>	50	36	14
<b>South Carolina</b>	27	30	3
<b>South Dakota</b>	45	46	1
<b>Tennessee</b>	21	21	0
<b>Texas</b>	4	2	2
<b>Utah</b>	27	22	5

### Algorithmen-Aversion im Controlling

<b>Staat</b>	<b>Forecast</b>	<b>Tatsächlicher Rang</b>	<b>Forecastinggenauigkeit</b>
<b>Vermont</b>	49	47	2
<b>Virginia</b>	9	10	1
<b>Washington</b>	18	13	5
<b>West Virginia</b>	49	49	0
<b>Wisconsin</b>	24	25	1
<b>Wyoming</b>	45	48	3

*Anhang 3: Forecastinggenauigkeit in der APDR*

<b>Forecastinggenauigkeit</b>	<b>Anzahl an Antworten</b>	<b>Kumulierte Häufigkeit</b>
<b>0</b>	12	5,7
<b>1</b>	16	13,4
<b>2</b>	16	21,1
<b>3</b>	12	26,8
<b>4</b>	12	32,5
<b>5</b>	10	37,3
<b>6</b>	8	41,1
<b>7 und schlechter</b>	123	100

*Anhang 4: Forecastinggenauigkeit in der APDB*

<b>Forecastinggenauigkeit</b>	<b>Anzahl an Antworten</b>	<b>Kumulierte Häufigkeit</b>
<b>0</b>	13	6,3
<b>1</b>	21	16,3
<b>2</b>	21	26,4
<b>3</b>	15	33,7
<b>4</b>	4	35,6
<b>5</b>	13	41,8
<b>6</b>	15	49,0
<b>7 und schlechter</b>	106	100

## Algorithmen-Aversion im Controlling

### Anhang 5: Experteninterviews

Interview	Position	Branche	Datum
A	Controller	Automobilindustrie (Zulieferer)	02.02.2022
B	Leiter Reporting	Maschinenbau	02.02.2022
C	Leiter Digitalisierung Rechtsprozesse	Automobilindustrie (OEM)	03.02.2022
D	Business Development	Maschinenbau	04.02.2022
E	Divers, Diskussion im ICV Arbeitskreis Stuttgart mit elf Teilnehmern	Divers	17.03.2022
F	Manager Planungs- und Forecastingsoftware	Softwareindustrie	17.03.2022
G	Projektmanager Unternehmensprozesse und Digitalisierung	Antriebstechnik	18.03.2022

### Anhang 6: Ablauf der Experimente

Zuordnung:

- 1: APC
- 1.1: APTP
- 1.2: APDB
- 1.3: APDR
- 2: AAC / CA / CBP / CGP
- 2.1: AATP
- 2.2: AADB
- 2.3: AADR
- 2.4: PA
- 2.5: NBP / NGP

Screen 1 (Alle)

We are researchers at the University of Ulm and we are investigating judgement and decision making. You will make judgements and predictions in the following study. Please be assured that your responses will be kept completely confidential.

In this survey, you will view information about states and make judgements about the amount of air traffic that departed from each state in 2011.

The study should take you approximately 10 minutes to complete.

Your participation in this research is voluntary. You have the right to withdraw at any point during the study, for any reason, and without prejudice.

By clicking the next button, you acknowledge that your participation in the study is voluntary and that you are aware that you may choose to terminate your participation in the study at any time and for any reason without penalty.

**Please enter your Mturk ID in the space below.**

Screen 2 (Alle)

**MTurk Experience**  
Your experiences on Mturk will be important for the following survey. Please indicate how often you participate in Mturk surveys in the question included below. In order to demonstrate that you have read these instructions, please select other and type the word dolphin as your answer to the question below. If you fail to do so you will not be able to take the survey. Please answer the question as honestly as you can.

Thank you for your participation.

**How often do you participate in Mturk surveys?**

- Never
- Less than once a month
- 1-3 times a month
- Once a week
- 2-3 times a week
- 4-5 times a week
- More than 5 times a week
- Other

Screen 3 (Alle)

Your answer to the previous question indicates that you did not read instructions carefully enough. You will not be able to participate in this survey.

Screen 4 (Alle)

**In this task you will estimate the rank of 1 U.S. state in terms of the number of airline passengers who departed from that state in 2011.**

Have you ever participated in a survey with this task?

- Yes, I have participated in this task before
- No, I have not participated in this task before

# Algorithmen-Aversion im Controlling

## Screen 5 (Alle)

## Screen 6 (1, 1.1, 1.3, 2, 2.1, 2.3, 2.4, 2.5)

Thank you for answering this question honestly. Since you have taken a similar survey before you do not qualify for this one.

In this study, **you will have the opportunity to earn additional money** by making accurate judgements. Please read all the information very carefully.

## Screen 7 (Alle)

## Screen 8 (Alle)

In this task you will estimate the rank of 1 U.S. state in terms of the number of airline passengers who departed from that state in 2011.

The state with the largest number of departing passengers is ranked 1st and the state with the smallest number of departing passengers is ranked 50th.

All of the data shown in this study represent real data from official government records.

You will make official estimates for 1 state. You will base your evaluation on a list of information about the state. Not all of the provided pieces of information are necessarily important indicators of passenger departures. You will receive the following information:

### Number of Major Airports

The number of major airports in the state as determined by the Bureau of Transportation. All states have smaller airports that this number does not account for

### Census Population Rank - 2010

The state's rank in terms of population in 2010 from the U.S. Census Bureau (1 = most populated U.S. state; 50 = least populated U.S. state)

### Number of Counties Rank

The state's rank in terms of its number of counties (1 = U.S. state with the most number of counties; 50 = U.S. state with the least number of counties)

### Median Household Income Rank - 2008

The state's rank in terms of median household income in 2008 from the U.S. Census Bureau (1 = U.S. state with the highest median income; 50 = U.S. state with the lowest median income)

### Domestic Travel Expenditure Rank - 2009

The state's rank in terms of money spent by U.S. citizens traveling to the state in 2009 from the U.S. travel association (1 = U.S. state with the most incoming expenditures; 50 = U.S. state with the least incoming expenditures)

## Screen 9 (Alle)

## Screen 10 (2.3)

In addition to information about the state, you will receive a prediction from a statistical model developed by transportation analysts. This model is designed to forecast each state's rank in terms of the number of passengers departing from that state in 2011. The model is based on data from 2006 through 2010, and it uses the same information that you will receive. The model does not have any additional information that you will not receive. This is a sophisticated model, put together by thoughtful analysts.

The statistical model predicts each state's rank in terms of the number of passengers that took off from that state in 2011. Scores range from 1 to 50. The model assigns lower ranks to states with more departing passengers.

**The model can only process 4 input factors at a time.**

You determine the 4 factors which you think the model should use to make the best forecast. We still show you all 5 input factors for your forecast.

Please choose 4 items that the model will process

- Number of Major Airports
- Census Population Rank – 2010
- Number of Counties Rank
- Median Household Income Rank – 2008
- Domestic Travel Expenditure Rank – 2009

## Screen 11 (2.2)

## Screen 12 (1.1, 2.1)

We want you to **do your best in this task** to get the best estimation.

Therefore you will go through 10 practice rounds to get experience with the data.

You will only have a limited amount of time available for your estimation. **Your estimation needs to be typed in within 12 seconds.**

## Screen 13 (2.4, 2.5)

## Screen 14 (2, 2.1, 2.3, 2.4, 2.5)

We have already conducted this survey a year ago. To help you assessing your performance, we will provide the performance data of other MTurkers to you. To evaluate your progress we provide you with this assessment after the first 5 practice estimates and after all 10 practice estimates. Your performance can either be in the best, middle oder worst third compared to other MTurkers performance.

Next, you will go through 10 practice estimates to gain experience with the data.

You will estimate the rank of each state in terms of how many airline passengers departed from that state in 2011. You will also see the statistical model's estimated rank of each state. Then, you will receive feedback indicating how close you estimated rank and the model's estimated rank were to each state's true rank.

[Next](#)

## Screen 15 (2, 2.1, 2.3)

After you complete the 10 practice estimates, you will go through 1 official estimate in which you will be compensated for your performance. The U.S. states you see in the practice round will be different from the one you see in the official round.

Before the official round, you will be given a choice of having your estimated rank or the model's estimated rank count as the official estimate.

During the official round, you will receive additional bonus money based on the accuracy of the official estimate. You can earn \$0 to \$1 depending on how close the official estimate is to the actual rank.

The bonus will be determined as follows:

- 1.00 - perfectly predict state's actual rank
- 0.85 - within 1 rank of state's actual rank
- 0.70 - within 2 ranks of state's actual rank
- 0.55 - within 3 ranks of state's actual rank
- 0.40 - within 4 ranks of state's actual rank
- 0.25 - within 5 ranks of state's actual rank
- 0.10 - within 6 ranks of state's actual rank

Please type the underlined sentences in the paragraph above this question in the text box.

# Algorithmen-Aversion im Controlling

## Screen 16a-25a (2.1)

Remaining Time: XX

Number of Major Airports	XX
Census Population Rank - 2010	XX
Number of Counties Rank	XX
Median Household Income Rank - 2008	XX
Domestic Travel Expenditure Rank - 2009	XX

What do you think this state's rank is in terms of departing airline passengers in 2011? (Please enter a number 1-50)

## Screen 16b-25b (2.1)

Your Prediction:

Model's Prediction:

State's Actual Rank:

## Screen 26a-36a außer 31 (2, 2.2, 2.3, 2.4, 2.5)

Number of Major Airports	XX
Census Population Rank - 2010	XX
Number of Counties Rank	XX
Median Household Income Rank - 2008	XX
Domestic Travel Expenditure Rank - 2009	XX

What do you think this state's rank is in terms of departing airline passengers in 2011? (Please enter a number 1-50)

## Screen 26b-36b außer 31 (2, 2.2, 2.3, 2.4, 2.5)

Your Prediction:

Model's Prediction:

State's Actual Rank:

## Screen 31a (2.4)

Performance evaluation on your first 5 practice estimates:

We also conducted this survey a year ago. At that time, we recorded the needed time and the performance of MTurkers incentivized estimate.

For their incentivized estimate, the other MTurkers could choose to have either their own estimated rank or the statistical model's estimated rank determine their bonus. In other words, they chose to be paid based on their accuracy, or chose to be paid based on the model's accuracy.

We also measured your needed time and performance in the first 5 practice estimates.

Compared to the other MTurkers with a similar amount of time needed for their incentivized forecast, your accuracy is within the

top 33 %

This means, that you performed

better than at least 66 %

of the MTurkers who have done this survey before.

Next

## Screen 31b (2.5)

Performance evaluation on your first 5 practice estimates:

We also conducted this survey a year ago. At that time, we recorded the needed time and the performance of MTurkers incentivized estimate.

For their incentivized estimate, the other MTurkers could choose to have either their own estimated rank or the statistical model's estimated rank determine their bonus. In other words, they chose to be paid based on their accuracy, or chose to be paid based on the model's accuracy.

We also measured your needed time and performance in the first 5 practice estimates.

Compared to the other MTurkers with a similar amount of time needed for their incentivized forecast, your accuracy is within the

worst 33 %

This means, that you performed

worse than at least 66 %

of the MTurkers who have done this survey before.

Next

## Screen 37a (2.4)

Performance evaluation on your 10 practice estimates:

For their incentivized estimate, the other MTurkers could choose to have either their own estimated rank or the statistical model's estimated rank determine their bonus. In other words, they chose to be paid based on their accuracy, or chose to be paid based on the model's accuracy.

We also measured your needed time and performance within all 10 practice estimates.

Compared to the other MTurkers with a similar amount of time needed for their incentivized forecast, your accuracy now is within the

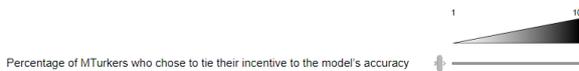
top 33 %

This means, that you performed

better than at least 66 %

of the MTurkers who have done this survey before.

What do you think, how many of the other MTurkers chose to tie their incentive to the model's accuracy?



Next

## Screen 37b (2.5)

Performance evaluation on your 10 practice estimates:

For their incentivized estimate, the other MTurkers could choose to have either their own estimated rank or the statistical model's estimated rank determine their bonus. In other words, they chose to be paid based on their accuracy, or chose to be paid based on the model's accuracy.

We also measured your needed time and performance within all 10 practice estimates.

Compared to the other MTurkers with a similar amount of time needed for their incentivized forecast, your accuracy is within the

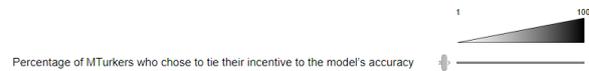
worst 33 %

This means, that you performed

worse than at least 66 %

of the MTurkers who have done this survey before.

What do you think, how many of the other MTurkers chose to tie their incentive to the model's accuracy?



Next

# Algorithmen-Aversion im Controlling

## Screen 38 (1.2)

Next, you will go through 1 official estimate.

You will be given a choice of having your estimated rank or the model's estimated rank count as the official estimate.

Your goal is to do your best to get the best estimation in this task.

Please type the underlined sentence in the paragraph above this question in the text box.

## Screen 39 (1, 1.1, 1.3)

Next, you will go through 1 official estimate.

You will be given a choice of having your estimated rank or the model's estimated rank count as the official estimate.

During the official round, you will receive additional bonus money based on the accuracy of the official estimate. You can earn \$0 to \$1 depending on how close the official estimate is to the actual rank.

The bonus will be determined as follows:

- 1.00 - perfectly predict state's actual rank
- 0.85 - within 1 rank of state's actual rank
- 0.70 - within 2 rank of state's actual rank
- 0.55 - within 3 rank of state's actual rank
- 0.40 - within 4 rank of state's actual rank
- 0.25 - within 5 rank of state's actual rank
- 0.10 - within 6 rank of state's actual rank

Please type the underlined sentences in the paragraph above this question in the text box.

## Screen 40 (1, 1.1, 1.2, 1.3)

You will receive the following information:

### Number of Major Airports

The number of major airports in the state as determined by the Bureau of Transportation. All states have smaller airports that this number does not account for

### Census Population Rank - 2010

The state's rank in terms of population in 2010 from the U.S. Census Bureau (1 = most populated U.S. state; 50 = least populated U.S. state)

### Number of Counties Rank

The state's rank in terms of its number of counties (1 = U.S. state with the most number of counties; 50 = U.S. state with the least number of counties)

### Median Household Income Rank - 2008

The state's rank in terms of median household income in 2008 from the U.S. Census Bureau (1 = U.S. state with the highest median income; 50 = U.S. state with the lowest median income)

### Domestic Travel Expenditure Rank - 2009

The state's rank in terms of money spent by U.S. citizens traveling to the state in 2009 from the U.S. travel association (1 = U.S. state with the most incoming expenditures; 50 = U.S. state with the least incoming expenditures)

## Screen 41 (2.2)

Next, you will go through 1 official estimate.

You will receive the same information that you received in the training rounds:

### Number of Major Airports

The number of major airports in the state as determined by the Bureau of Transportation. All states have smaller airports that this number does not account for

### Census Population Rank - 2010

The state's rank in terms of population in 2010 from the U.S. Census Bureau (1 = most populated U.S. state; 50 = least populated U.S. state)

### Number of Counties Rank

The state's rank in terms of its number of counties (1 = U.S. state with the most number of counties; 50 = U.S. state with the least number of counties)

### Median Household Income Rank - 2008

The state's rank in terms of median household income in 2008 from the U.S. Census Bureau (1 = U.S. state with the highest median income; 50 = U.S. state with the lowest median income)

### Domestic Travel Expenditure Rank - 2009

The state's rank in terms of money spent by U.S. citizens traveling to the state in 2009 from the U.S. travel association (1 = U.S. state with the most incoming expenditures; 50 = U.S. state with the least incoming expenditures)

Your goal is to do your best to get the best estimation in this task.

Please type the underlined sentence in the paragraph above this question in the text box.

## Screen 42 (2, 2.1, 2.3, 2.4, 2.5)

Next, you will go through 1 official estimate.

You will receive the same information that you received in the training rounds:

### Number of Major Airports

The number of major airports in the state as determined by the Bureau of Transportation. All states have smaller airports that this number does not account for

### Census Population Rank - 2010

The state's rank in terms of population in 2010 from the U.S. Census Bureau (1 = most populated U.S. state; 50 = least populated U.S. state)

### Number of Counties Rank

The state's rank in terms of its number of counties (1 = U.S. state with the most number of counties; 50 = U.S. state with the least number of counties)

### Median Household Income Rank - 2008

The state's rank in terms of median household income in 2008 from the U.S. Census Bureau (1 = U.S. state with the highest median income; 50 = U.S. state with the lowest median income)

### Domestic Travel Expenditure Rank - 2009

The state's rank in terms of money spent by U.S. citizens traveling to the state in 2009 from the U.S. travel association (1 = U.S. state with the most incoming expenditures; 50 = U.S. state with the least incoming expenditures)

## Screen 43 (1, 2)

For the official estimate, you can choose to have either your estimated rank or the statistical model's estimated rank determine your bonus. In other words, you can choose to be paid based on your accuracy, or you can choose to be paid based on the model's accuracy. You will make an estimate no matter which option you choose.

The bonus will be determined as follows:

- 1.00 - perfectly predict state's actual rank
- 0.85 - within 1 rank of state's actual rank
- 0.70 - within 2 rank of state's actual rank
- 0.55 - within 3 rank of state's actual rank
- 0.40 - within 4 rank of state's actual rank
- 0.25 - within 5 rank of state's actual rank
- 0.10 - within 6 rank of state's actual rank

Would you like your estimated rank or the model's estimated rank to determine your bonus?

Use only the statistical model's estimated rank to determine my bonus.

Use only my estimated rank to determine my bonus.

## Screen 44 (1.3)

For the official estimate, you can choose to have either your estimated rank or the statistical model's estimated rank determine your bonus. In other words, you can choose to be paid based on your accuracy, or you can choose to be paid based on the model's accuracy. You will make an estimate no matter which option you choose.

The bonus will be determined as follows:

- 1.00 - perfectly predict state's actual rank
- 0.85 - within 1 rank of state's actual rank
- 0.70 - within 2 rank of state's actual rank
- 0.55 - within 3 rank of state's actual rank
- 0.40 - within 4 rank of state's actual rank
- 0.25 - within 5 rank of state's actual rank
- 0.10 - within 6 rank of state's actual rank

The model can only process 4 input factors at a time.

You determine the 4 factors which you think the model should use to make the best forecast. We still show you all 5 input factors for your forecast.

Please choose 4 items that the model will process

- Number of Major Airports
- Census Population Rank - 2010
- Number of Counties Rank
- Median Household Income Rank - 2008
- Domestic Travel Expenditure Rank - 2009

Would you like your estimated rank or the model's estimated rank to determine your bonus?

Use only my estimated rank to determine my bonus.

Use only the statistical model's estimated rank to determine my bonus.

You can make changes to you prior choice for the models 4 input factors. Do you wish to make changes to the 4 input factors for the model?

No

Yes

# Algorithmen-Aversion im Controlling

## Screen 46 (2.3)

For the official estimate, you can choose to have either your estimated rank or the statistical model's estimated rank determine your bonus. In other words, you can choose to be paid based on your accuracy, or you can choose to be paid based on the model's accuracy. You will make an estimate no matter which option you choose.

The bonus will be determined as follows:

- 1.00 - perfectly predict state's actual rank
- 0.85 - within 1 rank of state's actual rank
- 0.70 - within 2 rank of state's actual rank
- 0.55 - within 3 rank of state's actual rank
- 0.40 - within 4 rank of state's actual rank
- 0.25 - within 5 rank of state's actual rank
- 0.10 - within 6 rank of state's actual rank

**The model can only process 4 input factors at a time.**

You determine the 4 factors which you think the model should use to make the best forecast. We still show you all 5 input factors for your forecast.

Please choose 4 items that the model will process

- Number of Major Airports
- Census Population Rank – 2010
- Number of Counties Rank
- Median Household Income Rank – 2008
- Domestic Travel Expenditure Rank – 2009

**Would you like your estimated rank or the model's estimated rank to determine your bonus?**

Use only the statistical model's estimated rank to determine my bonus.

Use only my estimated rank to determine my bonus.

## Screen 47 (1.2, 2.2)

Please do your best job on this task to get the best estimation. You can choose to have either your estimated rank or the statistical model's estimated rank determine your official result. In other words, you can choose your accuracy, or you can choose the model's accuracy. You will make an estimate no matter which option you choose.

**Would you like your estimated rank or the model's estimated rank to determine your official round to get the best estimation?**

Use only the statistical model's estimated rank to get the best result.

Use only my estimated rank to get the best result.

## Screen 48 (1.1, 2.1)

For the official estimate, you can choose to have either your estimated rank or the statistical model's estimated rank determine your bonus. In other words, you can choose to be paid based on your accuracy, or you can choose to be paid based on the model's accuracy. You will make an estimate no matter which option you choose.

The bonus will be determined as follows:

- 1.00 - perfectly predict state's actual rank
- 0.85 - within 1 rank of state's actual rank
- 0.70 - within 2 rank of state's actual rank
- 0.55 - within 3 rank of state's actual rank
- 0.40 - within 4 rank of state's actual rank
- 0.25 - within 5 rank of state's actual rank
- 0.10 - within 6 rank of state's actual rank

**There will only be a limited amount of time available for your estimation. In the official round your estimation needs to be typed in within 12 seconds.**

**Would you like your estimated rank or the model's estimated rank to determine your bonus?**

Use only the statistical model's estimated rank to determine my bonus.

Use only my estimated rank to determine my bonus.

## Screen 49a und 49b (Alle, randomisierte Reihenfolge)

On average, how many ranks do you think the model's estimates are away from states' actual ranks?

An answer of zero would mean that you think the model perfectly estimates all of the ranks. An answer of one would mean that you think the model's estimates are off by 1 rank, on average.

Your answer can range from 0-50.

On average, how many ranks do you think your estimates are away from states' actual ranks?

An answer of zero would mean that you think you perfectly estimate all of the ranks. An answer of one would mean that you think your estimates are off by 1 rank, on average.

Your answer can range from 0-50.

## Screen 50a und 50b (Alle, randomisierte Reihenfolge)

**How much confidence do you have in your estimates?**

None  Little  Some  A Fair Amount  A Lot

**How much confidence do you have in the statistical model's estimates?**

None  Little  Some  A Fair Amount  A Lot

## Screen 51 (Alle)

**How likely is the model to make a really bad estimate?**

Extremely Unlikely  Very Unlikely  Unlikely  Somewhat Unlikely  Undecided  Somewhat Likely  Likely  Very Likely  Extremely Likely

Now, you will go through 1 official estimate.

## Screen 52 (1, 1.2, 1.3, 2, 2.1, 2.2, 2.3, 2.4, 2.5)

# Algorithmen-Aversion im Controlling

## Screen 53 (1.1)

Now, you will go through 1 official estimate.

For your estimate you will receive the data shown as below. As you only have a limited amount of time available for your estimation, please read through this very carefully and make sure you understand each category and get comfortable with the way the data is displayed.

The data will be shown as follows:

Remaining Time: XX

Number of Major Airports	XX
Census Population Rank - 2010	XX
Number of Counties Rank	XX
Median Household Income Rank - 2008	XX
Domestic Travel Expenditure Rank - 2009	XX

What do you think this state's rank is in terms of departing airline passengers in 2011? (Please enter a number 1-50)

By clicking on the next button your official estimation starts with 12 seconds time limit.

## Screen 54 a (1.1, 2.1)

Remaining Time: XX

Number of Major Airports	XX
Census Population Rank - 2010	XX
Number of Counties Rank	XX
Median Household Income Rank - 2008	XX
Domestic Travel Expenditure Rank - 2009	XX

What do you think this state's rank is in terms of departing airline passengers in 2011? (Please enter a number 1-50)

## Screen 54 b (1.1, 2.1)

Your Prediction: .

Model's Prediction:

Do you think that your estimate or the Model's estimate is closer to the last state's true rank?

My estimate

The model's estimate

## Screen 55 a (1, 1.2, 1.3, 2, 2.2, 2.3, 2.4, 2.5)

## Screen 55 b (1, 1.2, 1.3, 2, 2.2, 2.3, 2.4, 2.5)

Number of Major Airports	XX
Census Population Rank - 2010	XX
Number of Counties Rank	XX
Median Household Income Rank - 2008	XX
Domestic Travel Expenditure Rank - 2009	XX

What do you think this state's rank is in terms of departing airline passengers in 2011? (Please enter a number 1-50)

Your Prediction: .

Model's Prediction:

Do you think that your estimate or the Model's estimate is closer to the last state's true rank?

My estimate

The model's estimate

## Screen 56 (1, 1.1, 1.3, 2, 2.1, 2.3, 2.4, 2.5)

## Screen 57 (1.2, 2.2)

You will learn and receive your bonus after all HITs have been completed.

If you earned a bonus of \$0, we will give you a \$0.01 bonus.

You will receive your bonus of \$1 for doing your best after all HITs have been completed.

## Screen 58(Alle)

## Screen 59 (Alle)

What is your age?

[Please choose] ▾

What is your gender

- Male
- Female

What is the highest level of education you have completed?

- Less than High School
- High School / GED
- Some College
- 2-year College Degree
- 4-year College Degree
- Masters Degree
- Professional Degree (JD,MD)
- Doctoral Degree

Thanks for participating!

Please copy and paste the following Survey Code to Mechanical Turk to receive credit: 1987MJ1

Please do not discuss any of the content in this survey with other people, including the choices that you made. Discussing any content from our survey will invalidate the results.

We are very aware of your performance and we will pay your bonus as quickly as possible!

## Eigenständigkeitserklärung

Ich versichere, dass ich die vorliegende Arbeit mit dem Titel

*Algorithmen-Aversion im Controlling  
Experimente zur Wirkung des Arbeitsumfeldes  
auf die Algorithmen-Aversion in Forecasts*

selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe; aus fremden Quellen entnommene Passagen und Gedanken sind als solche kenntlich gemacht.

## Declaration of Authorship

I hereby certify that the dissertation entitled

*Algorithmen-Aversion im Controlling  
Experimente zur Wirkung des Arbeitsumfeldes  
auf die Algorithmen-Aversion in Forecasts*

is entirely my own work except where otherwise indicated. Passages and ideas from other sources have been clearly indicated.

Name

Markus Jung

---

Unterschrift / Signature

---

Datum / Date

Stuttgart, Juli 2022

---