
Zusatzkapitel 2: Statistische Modellierung und Testbarkeit des Modells der Frame-Selektion*

Um formale *theoretische* Modelle wie das Modell der Frame-Selektion (MFS) oder Rational-Choice-Theorien in der quantitativen Sozialforschung anwenden zu können, müssen diese in *statistische* Modelle übersetzt werden (Achen 2002). In dieser Arbeit werden *logistische Regressionsmodelle* verwendet. Bei allen Schwierigkeiten und Begrenzungen, die bei der Anwendung derartiger Regressionsmodelle zu beachten sind, eignen sie sich doch am besten für einen kritischen Test der Interaktionshypothesen des MFS. Insbesondere erlauben sie den hierarchischen Test einer Modellspezifikation, welche die aus dem MFS ableitbaren Interaktionseffekte enthält, gegen eine alternative Spezifikation ohne diese Interaktionseffekte.¹ Dieses Kapitel diskutiert die Übersetzung der MFS-Hypothesen in statistische Modellierungen und dabei auftretende Herausforderungen. Es bildet damit gleichsam das Verbindungsglied zwischen den Theorie-Kapiteln und den empirischen Analysen der Wahlteilnahme und der Rettung von Juden während des Zweiten Weltkrieges.

* Dieses Zusatzkapitel kann wie folgt zitiert werden: Kroneberg, Clemens (2011): Zusatzkapitel 2 zu „Die Erklärung sozialen Handelns“: Statistische Modellierung und Testbarkeit des Modells der Frame-Selektion. URL: <http://vs-verlag.de/tu/Kroneberg-Erklärung>. Stand: 01.06.2011

¹ Eine alternative statistische Methode, die zum Test von MFS-Hypothesen geeignet wäre, stellen sog. Boolean Regressionsmodelle dar (Braumoeller 2003, 2004). Sie vermögen über die Betrachtung hinreichender Bedingungen bereits von sich aus spezifischer die Zwei-Modi-Logik des MFS abzubilden.

1 Das binäre logistische Regressionsmodell als Zufallsnutzenmodell

In der Form sog. Discrete-Choice-Modelle existiert eine Reihe von Regressionsmodellen, die eine enge Übereinstimmung zwischen theoretischer und statistischer Modellierung ermöglichen und somit die Kluft zwischen Rational-Choice-Theorien und empirischen Daten zu überbrücken helfen. In den folgenden empirischen Anwendungen wird das binäre logistische Regressionsmodell verwendet (Maier und Weiss 1990; Manski 1995: 91ff.), das sich auf Basis eines Zufallsnutzenmodells herleiten lässt (vgl. etwa Long 1997; Nagler 1994).

Ausgangspunkt ist die Entscheidung zwischen zwei diskreten Alternativen. Ich nenne sie im Folgenden Alternative 0 und Alternative 1, wobei die Variable Y_i den Wert 0 bzw. 1 annimmt, je nachdem welche Alternative ein Individuum i wählt. In den bislang dargestellten formalen Handlungsmodellen – etwa der SEU-Theorie oder dem MFS – wird implizit davon ausgegangen, dass die Entscheidung eines Akteurs auf Basis des Modells deterministisch vorhergesagt werden kann. Praktisch ist dies jedoch unmöglich, da keine empirische Studie *alle* Determinanten des Entscheidungsverhaltens vollständig erheben kann (Manski 1995: 91f.). Die Gleichungen dieser theoretischen Modelle, die das Verhalten oder die Disposition für ein Verhalten als Funktion anderer Faktoren beschreiben, müssen daher um einen Term ε erweitert werden, der diesen Vorhersagefehler oder Zufallseinfluss wiedergibt.

Zur Illustration sei davon ausgegangen, dass der subjektive Erwartungsnutzen aus der Wahl der Alternative 1 von drei Faktoren abhängt: Der Höhe der Kosten C , unter die hier vereinfachend auch die Opportunitätskosten (d.h. $SEU(\text{Alternative } 0)$) subsumiert werden, der Höhe des Nutzens U aus bestimmten positiv bewerteten Folgen sowie deren subjektiver Eintrittswahrscheinlichkeit p . Dies entspricht folgender SEU-Gleichung:

$$SEU(\text{Alternative } 1) = pU - C + \varepsilon \quad (7.1)$$

Die Konstrukte $SEU(\text{Alternative } 1)$, pU und C sind sog. latente Variablen, die nicht direkt empirisch beobachtet werden können. Im Rahmen einer direkten Teststrategie versucht man sie nichtsdestotrotz so gut wie möglich durch Indikatoren zu messen (vgl. Brüderl 2004). Entsprechend erhält man

für jede Person i einen Vektor X_i , der ihre Messwerte auf diesen erklärenden Variablen enthält. Regressionsanalytisch geht es vor allem darum, das relative Gewicht dieser Variablen, also ihre Einflussstärke zu schätzen. Der Vektor der Regressionsgewichte wird mit β notiert. Die abhängige Variable, auf die sich diese Erwartungs-, Kosten- und Nutzenterme beziehen, ist zunächst das SEU-Gewicht der Alternative 1. Im Kontext der statistischen Modellierung bezeichnet man diese Variable allgemein als Y^* . Es ergibt sich somit das folgende statistische Modell:

$$Y_i^* = X_i\beta + \varepsilon_i \quad (7.2)$$

Um die Parameter β_j , also die Einflussstärke der unabhängigen Variablen schätzen zu können, muss der Erwartungsnutzen aus der Wahl der Alternative 1, Y^* , gemessen werden. Dieser ist jedoch ebenfalls eine latente Variable, deren Messung noch schwieriger erscheint. Überlicherweise wird stattdessen direkt die von einem Befragten getroffene oder beabsichtigte Entscheidung erhoben. Bei zwei Handlungsalternativen ist dies eine dichotome Variable, welche die Werte 0 und 1 annehmen kann. Der Zusammenhang zwischen ihr und der zu Grunde liegenden latenten Variablen Y^* ist theoretisch vorgegeben: Übersteigt der Nutzen aus der Wahl von Alternative 1 einen bestimmten Schwellenwert τ' , so wird sich ein Akteur für Alternative 1 entscheiden ($Y = 1$). Dies lässt sich formal schreiben als:

$$\begin{aligned} Y_i &= 1 && \text{wenn } Y_i^* = X_i\beta + \varepsilon_i > \tau' \\ Y_i &= 0 && \text{sonst.} \end{aligned}$$

Der Schwellenwert τ' entspricht einer Konstanten β_0 . In der regressionsanalytischen Notation geht er als zu schätzender Parameter in den Vektor β ein (und X_i enthält entsprechend für jeden Befragten ein zusätzliches Element mit konstantem Wert 1). Diese Redefinition von β und X_i vereinfacht die weitere Notation merklich, da τ' nicht separat mitgeführt zu werden braucht.

Auf Basis der dargestellten Beziehungen lässt sich die Wahrscheinlichkeit betrachten, dass sich ein Befragter i für Alternative 1 entscheidet $P_i(Y_i = 1)$. Diese Wahrscheinlichkeit hängt nach dem obigen Modell von den Werten des Befragten auf den unabhängigen Variablen, also von X_i ab:

$$P_i(Y_i = 1 | X_i) = \text{prob}(X_i\beta + \varepsilon_i > 0)$$

Dies lässt sich umformen zu:

$$P_i(Y_i = 1 | X_i) = \text{prob}(\varepsilon_i > -X_i\beta) = 1 - F(-X_i\beta), \quad (7.3)$$

wobei F die kumulative Dichtefunktion des Fehlerterms ε bezeichnet. Um das statistische Modell schätzen zu können, ist es erforderlich eine Annahme über die Verteilung dieser Zufallsvariablen in der Population zu treffen (Manski 1995: 93). Es ist gängige Praxis, anzunehmen, dass ε unabhängig von den Werten der X -Variablen verteilt ist und dass die Verteilung bis auf den Nullpunkt und die Skaleneinheit bekannt ist. Im logistischen Regressionsmodell wird angenommen, dass ε logistisch mit einem Mittelwert von null und einer Varianz von $\pi^2/3$ verteilt ist. Setzt man die sich so ergebende kumulative Dichtefunktion ein, so ergibt sich das bekannte logistische Regressionsmodell:

$$P_i(Y_i = 1 | X_i) = 1 - F(-X_i\beta) = 1 - \frac{\exp(-X_i\beta)}{1 + \exp(-X_i\beta)} = \frac{1}{1 + \exp(-X_i\beta)} = \frac{\exp(X_i\beta)}{1 + \exp(X_i\beta)}$$

Aus dem linearen Modell für die latente Variable „Erwartungsnutzen aus der Wahl der Alternative 1“ ergibt sich somit ein nicht-lineares Modell für die Wahrscheinlichkeit dieser Wahl bzw. für die Wahrscheinlichkeit, auf der manifesten Variablen Y den Wert 1 zu beobachten.

Um umgekehrt vom logistischen Regressionsmodell wieder zu der latenten Variablen zu gelangen, muss man sog. Logits (logarithmierte Odds) betrachten. Dann resultiert auf der rechten Seite der Regressionsgleichung wieder eine Linearkombination der unabhängigen Variablen, die den linearen Effekten im ursprünglichen theoretischen Modell entspricht:

$$Y_i^* = \ln \left(\frac{P_i(Y_i = 1 | x_i)}{P_i(Y_i = 0 | x_i)} \right) = X_i\beta$$

Im Vergleich zu der nicht-linearen Wirkung der Prädiktoren auf die Wahrscheinlichkeit ergibt sich hier eine *theoretisch* unmittelbar bedeutsame Inter-

pretation der Regressionskoeffizienten: Diese stellen die Effekte der Prädiktoren auf die unbeobachtete latente Variable dar (Long 1997: 40ff.).

Eine weitere Interpretationsmöglichkeit erhält man, wenn man die Logits entlogarithmiert. Dann resultiert die Formulierung mit sog. Odds-Ratios (Long 1997: 79ff.). Die Odds entsprechen dem Verhältnis der Wahrscheinlichkeiten von Alternative 1 und Alternative 0:

$$\text{Odds}_i = \frac{P_i(Y_i = 1 | x_i)}{P_i(Y_i = 0 | x_i)} = \frac{P_i(Y_i = 1 | x_i)}{1 - P_i(Y_i = 1 | x_i)} = \exp \beta X_i$$

Wie im deutschsprachigen Raum üblich (siehe z.B. Kohler und Kreuter 2008: 263ff.), bezeichne ich die Odds als *Chance*, mit der es zur Wahl von Alternative 1 kommt. Beispielsweise beträgt die Chance 3 zu 1 bei einer Wahrscheinlichkeit von 0.75 für $Y_i = 1$ und einer Wahrscheinlichkeit von 0.25 für $Y_i = 0$. Das *Odds-Ratio* (Chancenverhältnis, kurz: OR) gibt an, um welchen multiplikativen Faktor sich diese Chance für eine gegebene Veränderung einer unabhängigen Variablen erhöht oder verringert. Beispielsweise entspricht eine 3-zu-1-Chance im Vergleich zu einer 1-1-Chance (beide Wahrscheinlichkeiten gleich 0.50) einer 3-fach höheren Chance für $Y_i = 1$.

Im folgenden Kapitel wird gezeigt, wie sich im Rahmen des logistischen Regressionsmodells die spezifischen Interaktionshypothesen des MFS testen lassen. Wie sich zeigt, ist dies nicht im Rahmen der Wahrscheinlichkeitsformulierung möglich, sondern erfordert die Betrachtung von Logit-Effekten oder Odds-Ratios.

2 Die Interaktionshypothesen des MFS im logistischen Regressionsmodell: Modellinhärente versus variablenspezifische Interaktionseffekte

Im Folgenden beschränke ich mich auf eine der in Kapitel 6.5 abgeleiteten Interaktionshypothesen des MFS. Unter der Annahme, dass ein Akteur in einer hinreichend definierten Situation ein Skript aktiviert hat, das die Handlungswahl hinreichend regelt, lässt Hypothese 1.3 Folgendes erwarten: Je stärker dieses Skript mental verankert ist, desto geringer ist tendenziell der Einfluss anderer Anreize und Alternativen auf die Handlungsselektion. Zu-

dem gilt generell, dass kein signifikanter Einfluss dieser Anreize und Alternativen mehr feststellbar sein sollte, wenn das Aktivierungsgewicht bzw. die Skript-Verankerung maximal ist (siehe Hypothese 1). Die anderen Anreize und Alternativen sind nur im rc-Modus verhaltensrelevant, wohingegen ein Akteur im as-Modus *unabhängig* von diesen Faktoren einem Skript folgt. Im Folgenden gehe ich davon aus, dass es sich bei dem Skript um eine Norm handelt.

Rational-Choice-Theorien betrachten internalisierte Normen dagegen als einen Anreiz unter anderen. Im betrachteten SEU-Gewicht lässt sich ihre Bedeutung entsprechend als zusätzlicher Nutzenterm berücksichtigen: Aus der Wahl der Alternative 1 entsteht ein Konsumnutzen U_{Norm} , falls und insoweit der Akteur diese Alternative als normativ geboten empfindet. Die obige SEU-Gleichung wird also einfach um diesen Faktor erweitert:

$$\text{SEU}(\text{Alternative 1}) = pU - C + U_{\text{Norm}} + \varepsilon \quad (7.4)$$

Das MFS behauptet dagegen die Irrelevanz von pU und C bei Entscheidungen im as-Modus. Wenn die Norm stark genug verinnerlicht ist und die weiteren, oben genannten Bedingungen vorliegen, entscheidet sich der Akteur unabhängig von diesen Faktoren für Alternative 1. Entsprechend gilt:

$$\text{Disposition zur Wahl von Alternative 1} = \begin{cases} \infty & , \text{ wenn } a_k \geq 1 - C/U_{\text{rc}} \\ \text{SEU}(\text{Alternative 1}) & , \text{ wenn } a_k < 1 - C/U_{\text{rc}} \end{cases}$$

Diese Kontrastierung bezieht sich auf die *Disposition*, die Alternative 1 zu wählen, also Y^* . Die SEU-Gleichung geht diesbezüglich von einem *additiven* Effekt der Norm aus. Das MFS dagegen sagt vorher, dass sie die Wirkung anderer Anreizvariablen *moderiert*.

Beide Theorien implizieren somit unterschiedliche Spezifikationen der obigen Regressionsgleichung 7.2. Vereinfachend sei neben einer Normvariablen X_N nur eine Anreizvariable X_A betrachtet (wobei Ausprägungen mit x_N bzw. x_A bezeichnet werden). Die einfache SEU-Modellierung geht davon aus, dass in der Population gilt:

$$Y^* = \beta_0 + \beta_N x_N + \beta_A x_A$$

Das MFS sagt dagegen einen Interaktionseffekt zwischen diesen beiden Variablen vorher, der sich – wie auch in der linearen Regression üblich – mittels eines Produktterms modellieren lässt:

$$Y^* = \beta_0 + \beta_N x_N + \beta_A x_A + \beta_{IE} x_N x_A = \beta_0 + \beta_N x_N + (\beta_A + \beta_{IE} x_N) x_A$$

Der Parameter des Produktterms, β_{IE} , gibt an, wie der Einfluss der Anreizvariablen mit der Ausprägung der Normvariablen variiert (und umgekehrt). Das MFS sagt erstens vorher, dass dieser Parameter von seinem Vorzeichen her dem Effekt der Anreizvariablen β_A entgegenwirkt. Der zweite Teil der betrachteten Hypothese impliziert, dass dieser Interaktionseffekt so stark ist, dass sich bei maximaler Ausprägung der Normvariablen kein Anreizeffekt mehr ergibt, dass also der konditionale Anreizeffekt ($\beta_A + \beta_{IE} x_N$) null ist. Die betrachtete SEU-Modellierung lässt dagegen erwarten, dass der Parameter β_{IE} null ist.

Letzteres gilt jedenfalls, sofern man von einheitlichen Nutzenfunktionen ausgeht. Selbstverständlich könnte man im RC-Ansatz prinzipiell annehmen, dass es zwei Kategorien von Akteuren gibt: In die Nutzenfunktion der Akteure der Kategorie „as“ geht ausschließlich die betrachtete Norm ein, wohingegen Akteure der Kategorie „rc“ zusätzliche Anreize berücksichtigen. Aufgrund dieser prinzipiellen Möglichkeit gilt die Kontrastierung zwischen MFS und RC-Theorien nur mit der Einschränkung, dass letztere von einheitlichen, sich höchstens zufällig unterscheidenden Nutzenfunktionen ausgehen. Diese wird im Folgenden zu Grunde gelegt, da eine ad-hoc Unterscheidung zweier Akteurskategorien die in Kapitel 3 identifizierten, handlungstheoretischen Defizite des RC-Ansatzes offensichtlich nicht zu beheben vermag.

In Bezug auf die Disposition Y^* ist es also möglich, die konkurrierenden Hypothesen des MFS und der betrachteten SEU-Modellierung *theoretisch* zu unterscheiden und einen entsprechenden statistischen Test zu konzipieren. Wie oben dargestellt, lässt sich diese Dispositionsvariable auch im binären logistischen Regressionsmodell betrachten, wenn man die Logit-Formulierung verwendet. Dies ändert freilich nichts daran, dass die *kontinuierliche* Disposition zur Ausführung eines Verhaltens eine latente, also unbeobachtete Variable bleibt. Als empirisch gemessene Größe geht in das binäre logistische Regressionsmodell nur eine *dichotome* Intentions- oder Verhaltens-

variable ein. Betrachtet man eine derartige dichotome Verhaltensvariable Y, so scheint der Unterschied zwischen der MFS- und der SEU-Modellierung verloren zu gehen.

Auch die SEU-Gleichung 7.4, die Normen lediglich als Anreiz unter anderen berücksichtigt, prognostiziert nämlich für den Fall einer stark verinnerlichten Norm eine „unbedingte“ Befolgung von Alternative 1: Wenn der Konsumnutzen aus der Normbefolgung größer als ihre Kosten ist, wenn also $U_{\text{Norm}} > C$ gilt, wird sich ein Akteur für Alternative 1 entscheiden, und zwar unabhängig von der Höhe des anderen Anreizes pU. Es handelt sich hierbei ebenfalls um eine Form der statistischen Interaktion, da die Wirkung einer Variablen (pU) unterschiedlich groß ist, je nachdem welche Ausprägung eine andere Variable (U_{Norm}) aufweist.

Diese Form der Interaktion ergibt sich allein aus dem Übergang zur Betrachtung der *dichotomen* Verhaltensvariablen. Diese sog. *modellinhärenten* Interaktionseffekte bestehen selbst zwischen Variablen, von denen in Bezug auf die *Disposition* für ein Verhalten theoretisch angenommen wird, dass sie unabhängig voneinander, also rein additiv wirken. Derartige Interaktionen, wie sie nicht-linearen Regressionsmodellen inhärent sind, gilt es von den spezifischen Interaktionseffekten abzugrenzen, welche vom MFS vorhergesagt werden. Letztere sind sog. *variablenspezifische* Interaktionseffekte (Nagler 1994).

Die *modellinhärenten* Interaktionseffekte im logistischen Regressionsmodell beziehen sich auf die Wahrscheinlichkeit, die Alternative 1 zu wählen ($Y_i = 1$). Dass das statistische Modell diesbezüglich Interaktionen zwischen allen unabhängigen Variablen impliziert, wird deutlich, wenn man betrachtet, wie sich diese Wahrscheinlichkeit bei einer marginalen Änderung der k-ten X-Variablen verändert (Nagler 1994: 232). Dies entspricht der ersten Ableitung der im vorherigen Abschnitt aufgeführten Gleichung 7.3:

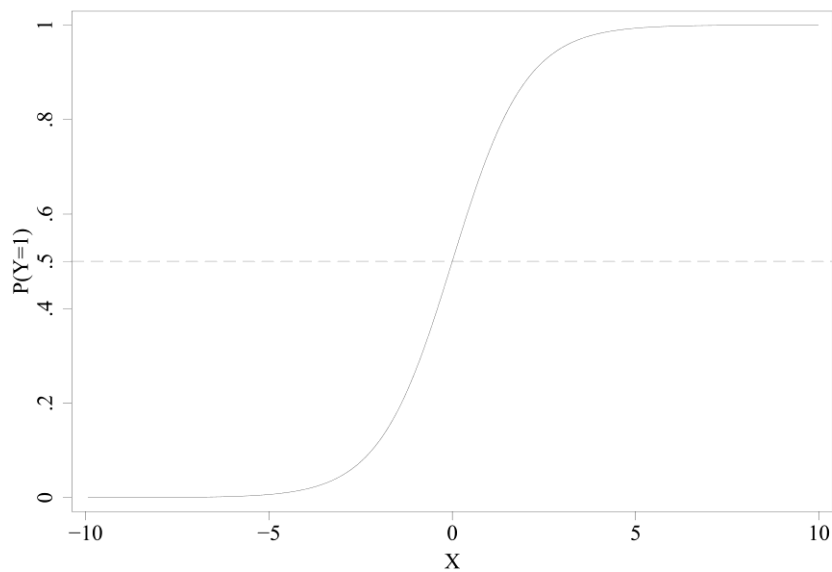
$$\frac{\partial P_i(Y_i = 1 | X_i)}{\partial x_k} = \frac{\partial [1 - F(-X_i\beta)]}{\partial x_k} = f(-X_i\beta)\beta_k \quad (7.5)$$

Die Veränderung der Wahrscheinlichkeit, die Alternative 1 zu wählen, hängt demnach nicht nur von dem Regressionsgewicht der jeweiligen Variablen (β_k) ab, sondern *zusätzlich* auch von dem Wert von $-X_i\beta$, also von den Wer-

ten der anderen unabhängigen Variablen (genauer: von der marginalen Dichtefunktion $f(\cdot)$ über diese Argumente).

Aufgrund der Symmetrie der angenommenen logistischen Verteilung ist der Effekt einer unabhängigen Variablen am stärksten, wenn die Ausprägungskombination der anderen unabhängigen Variablen zu einer Wahrscheinlichkeit von $P_i(Y_i = 1 \mid X_{i,-k}) = P_i(Y_i = 0 \mid X_{i,-k}) = 0.5$ führt. Dies ist unmittelbar einleuchtend, da in diesem Fall der Wert der k -ten unabhängigen Variablen ausschlaggebend dafür ist, ob die Alternative 1 oder die Alternative 0 wahrscheinlicher ist. Allgemein entspricht die Höhe des Effektes einer unabhängigen Variablen der *Steigung* der in Abbildung 1 dargestellten Funktion.

Abbildung 1: Nicht-Linearität der vorhergesagten Wahrscheinlichkeit im logistischen Regressionsmodell ($\beta_0 = 0, \beta_1 = 1$)



Variablenspezifische Interaktionseffekte, wie sie aus dem MFS abgeleitet werden können, gehen über derartige modellinhärente Interaktionen hinaus. Sie ergeben sich nicht erst mit dem Übergang zur Betrachtung des Verhaltens selbst bzw. der Wahrscheinlichkeit für das Verhalten, sondern werden be-

reits theoretisch in Bezug auf die metrische Dispositionsvariable angenommen. Da diese Interaktionen nur für *bestimmte* Variablen vorhergesagt werden, macht es Sinn von variablenspezifischen Interaktionseffekten zu sprechen – im Unterschied zu den modellinhärenten Interaktionseffekten, die sich bereits allein aus der funktionalen Form eines nicht-linearen Modells ergeben (Nagler 1994: 249). Wie bereits dargestellt, werden sie deshalb im Regressionsmodell durch den Einschluss eines zusätzlichen multiplikativen Terms $x_{jk} = x_j \times x_k$ gesondert spezifiziert. Leitet man die Wahrscheinlichkeit für die Wahl der Alternative 1 nach einer der beiden multiplikativ verknüpften Variablen ab, z.B. nach x_k , so erhält man

$$\frac{\partial P_i(Y_i = 1 | X_i)}{\partial x_k} = f(-X_i\beta)(\beta_k + \beta_{jk}x_{ji}). \quad (7.6)$$

Die marginale Veränderung der Wahrscheinlichkeit hängt also vom Wert von x_{jk} *über* dessen Einfluss auf die Summe $-X_i\beta$ *hinaus* ab. Die Ableitungen in den Gleichungen 7.5 und 7.6 unterscheiden sich genau hinsichtlich dieses zusätzlichen Einflusses.

Selbst wenn man die Wahrscheinlichkeitsinterpretation zu Grunde legt, kann also im Rahmen des logistischen Regressionsmodells *prinzipiell* zwischen den Interaktionshypothesen des MFS und konkurrierenden Hypothesen, die das Vorhandensein derartiger Interaktionen verneinen, diskriminiert werden. Trotz ihrer intuitiven Anschaulichkeit wird aber von der Wahrscheinlichkeitsinterpretation in den empirischen Anwendungen abgesehen. Denn wie Gleichung 7.6 zeigt, gehen in die Wahrscheinlichkeitswerte sowohl modellinhärente als auch variablenspezifische Interaktionseffekte ein. Um nur das Vorhandensein der letzteren statistisch zu testen, sollten Logit-Effekte oder Odds-Ratios betrachtet werden, da diese die theoretisch interessierenden variablenspezifischen Interaktionseffekte direkt zu isolieren und inferenzstatistisch zu bewerten erlauben.²

² Auf keinen Fall sollte man dagegen der von Ai und Norton vorgeschlagenen Interpretationsmethode für Interaktionseffekte in nicht-linearen Modellen folgen (Ai und Norton 2003). Diese definieren Interaktionseffekte rein technisch als Kreuzableitung des Erwartungswertes von Y und betrachten als Erwartungswert die Wahrscheinlichkeit, dass $Y = 1$. Diesen Interaktionseffekt erhält man, indem man die obige Gleichung 7.6 nochmals nach der anderen Komponentenvariablen (x_j) ableitet (Ai und Norton 2003: 124). Es re-

Die Kontrastierung konkurrierender Hypothesen auf der Ebene der theoretischen Modelle lässt sich also auch statistisch im Rahmen des binären Regressionsmodells konzeptionell aufrecht erhalten. Den Schlüssel dazu bildet die strikte Unterscheidung von modellinhärenten und variablenspezifischen Interaktionseffekten. Aus dem Übergang von der kontinuierlichen Verhaltensdisposition zur Betrachtung *dichotomer* Verhaltens- oder Intentionsvariablen resultiert allerdings noch ein zweites, praktisches Problem, dessen Bedeutung für die folgenden empirischen Anwendungen kaum überschätzt werden kann. Dieses bezieht sich auf die Aussagekraft inferenzstatistischer Tests der MFS-Hypothesen und steht im Mittelpunkt des folgenden Unterkapitels.

Zuvor sei jedoch noch kurz ein weiteres Problem diskutiert, das den Test von Interaktionshypothesen in logistischen Regressionsmodellen erschwert. Im Gegensatz zu linearen Regressionsmodellen können in logistischen Regressionsmodellen unterschiedliche Koeffizienten allein daraus resultieren, dass die abhängige Variable in einer Subgruppe besser vorhergesagt werden kann (geringere unbeobachtete Heterogenität) als in einer anderen (Allison 1999; Mood 2010). Dieses Problem besteht selbst dann, wenn die unberücksichtigten Variablen unabhängig von den berücksichtigten sind.

Statistisch signifikante Interaktionseffekte können daher nur in dem Maße als Evidenz für statistische Interaktionen angesehen werden, in dem man annimmt, dass die verglichenen Subgruppen die gleiche Fehlervarianz aufweisen (Allison 1999). Im Rahmen eines Tests der MFS-Hypothesen muss entsprechend angenommen werden, dass die spezifizierten Regressionsmodelle das Verhalten von Akteuren, die im as-Modus handeln, in etwa gleich gut vorhersagen wie das von Akteuren, die im rc-Modus handeln. Derartige Annahmen lassen sich jedoch selten rechtfertigen (Allison 1999; Mood 2010).

sultiert dann ein komplizierter Ausdruck, der modellinhärente und variablenspezifische Interaktionseffekte hoffnungslos vermischt. Für diesen „Interaktionseffekt“ gilt: „The interaction effect is always positive for some observations and negative for others.“ (Ai und Norton 2003: 128) Die Autoren äußern sich an keiner Stelle zur substantiellen Bedeutung dieses Phänomens. Tatsächlich ist es ein Artefakt, das sich direkt aus dem S-förmigen Verlauf der kumulativen logistischen bzw. Normalverteilung ergibt (siehe Abbildung 1). Theoretisch sinnvoll interpretierbare Ergebnisse erhält man dagegen nur, wenn man zwischen den beiden Arten von Interaktionseffekten systematisch trennt. Dies gilt jedenfalls, solange sich Theorien und Hypothesen nicht direkt auf die *Wahrscheinlichkeit* eines Ereignisses beziehen.

Es lohnt sich daher, kurz umgekehrt davon auszugehen, dass die in bisherigen Studien gefundenen MFS-Interaktionseffekte lediglich Artefakte unterschiedlich hoher Fehlervarianzen sind. In diesem Fall würden die beobachteten, *stärkeren* Anreizeffekte bei Befragten mit geringer Skript-Verankerung eine *geringere* statistische Fehlervarianz in dieser Subgruppe anzeigen (Allison 1999: 190; Mood 2010: 74). Aus zwei Gründen wäre theoretisch jedoch eher das Gegenteil zu erwarten, dass also die statistische Fehlervarianz bei Akteuren mit geringer Skript-Verankerung größer ist. Erstens dürften Entscheidungen im rc-Modus generell weniger vorhersehbar sein als spontane Selektionen im as-Modus. So ist von einer relativ großen Heterogenität hinsichtlich der Frage auszugehen, welche Anreize in Betracht gezogen werden und wie diese eingeschätzt werden. Zweitens lässt sich in Regressionsmodellen immer nur eine (sehr) begrenzte Anzahl dieser Anreizvariablen berücksichtigen.

Folgt man dieser Argumentation, so erscheint es unwahrscheinlich, dass signifikante, dem MFS entsprechende Interaktionseffekte in erster Linie Ausdruck unterschiedlich hoher Fehlervarianzen sind. Im Gegenteil: Dieses Problem logistischer Regressionsmodelle sollte die Bestätigung der betrachteten Interaktionshypothesen tendenziell eher erschweren.

3 Monte-Carlo-Simulationen zur statistischen Prüfbarkeit der MFS-Hypothesen im logistischen Regressionsmodell

Verglichen mit der theoretisch zu Grunde liegenden metrischen Dispositionsvariablen Y^* bedeutet der Übergang zur Betrachtung der *dichotomen* Verhaltensvariablen Y einen enormen Informationsverlust. Wie sich zeigen lässt, wird es dadurch außerordentlich schwierig, die Nullhypothese, dass die vom MFS vorhergesagten Interaktionseffekte *nicht* bestehen, zurückzuweisen.

Eine der hauptsächlichen Ursachen dieses Problems lässt sich bereits an der obigen Abbildung 1 veranschaulichen. Wie bereits erwähnt, sollte sich der stärkste Effekt einer Anreizvariablen für Fälle ergeben, deren „Ausgangswahrscheinlichkeit“ (Nagler 1994: 233) nahe an 0.5 liegt. Wenn aber bereits das Regressionsmodell *ohne* variablenspezifischen Interaktionseffekt sehr hohe Wahrscheinlichkeiten für die Wahl von Alternative 1 vorhersagt, tritt ein ceiling-Problem auf: In dem Bereich, in dem sich die Wahrschein-

lichkeit dem Wert 1 annähert, verbleibt nur noch wenig Spielraum für den Nachweis eines variablenspezifischen Interaktionseffektes – selbst wenn dieser tatsächlich bestünde.

Genau mit diesem Problem ist bei der Überprüfung der betrachteten MFS-Hypothese aber zu rechnen: Eine zentrale Moderatorvariable des MFS – der Internalisierungsgrad der Norm – hängt meist stark mit der Verhaltensvariablen zusammen. Bereits in einem Regressionsmodell ohne Produktterm führt eine starke Ausprägung dieser Variablen daher zu einer hohen Ausgangswahrscheinlichkeit und damit zu nur noch geringen marginalen Effekten anderer Anreizvariablen. Es ist daher schwierig, darüber hinaus noch einen variablenspezifischen Interaktionseffekt nachzuweisen.

Im Folgenden wird das Ausmaß dieses statistischen Problems mit Hilfe von Monte-Carlo-Simulationen abgeschätzt. Um diese Analysen zu verstehen, ist es sinnvoll, zunächst kurz den inferenzstatistischen Hintergrund in Erinnerung zu rufen.

In der Population kann entweder die MFS-Hypothese eines variablenspezifischen Interaktionseffektes wahr sein oder aber die Nullhypothese, der zufolge dieser Interaktionseffekt nicht besteht. Je nachdem welche Hypothese wahr ist, kann es in der statistischen Analyse einer Stichprobe zu zwei Arten von Fehlern kommen: Wenn man statistisch zu dem Schluss gelangt, dass die MFS-Hypothese wahr ist, obwohl sie tatsächlich falsch ist, begeht man den sog. α -Fehler (oder Fehler 1. Art). Um möglichst konservativ zu sein, möchte man die Wahrscheinlichkeit eines α -Fehler möglichst gering halten. Es ist bekanntlich Konvention, eine Nullhypothese nur dann zu verwerfen, wenn die Irrtumswahrscheinlichkeit weniger als fünf Prozent beträgt.

Der Fehler 2. Art ist der sog. β -Fehler. Er bezieht sich auf die Situation, in der man die Nullhypothese nicht zurückweist, obwohl die MFS-Hypothese tatsächlich wahr ist. Wenn ein variablenspezifischer Interaktionseffekt besteht, sollte ein statistisches Verfahren diesen mit einer möglichst hohen Wahrscheinlichkeit auch finden. Diese Wahrscheinlichkeit (die falsche Nullhypothese zu verwerfen) wird als *Teststärke* bezeichnet. Sie berechnet sich als $(1 - \beta)$, bildet also das Komplement zum β -Fehler.

In der Soziologie, wie auch in anderen Sozialwissenschaften, wird die Teststärke in empirischen Analysen häufig ignoriert (Cohen 1992). Oftmals wird aus der statistischen Insignifikanz eines Koeffizienten geschlossen, dass der Effekt offenbar nicht besteht (Aguinis 2004: 68f.). Dies ist insofern eine

Fehlinterpretation von Signifikanztests, als die Insignifikanz auch durch eine zu geringe Teststärke verursacht sein kann.

Gerade Tests von Interaktionseffekten weisen häufig eine ungenügende Teststärke auf (Aguinis 2004: 68f.). Auf der Basis umfassender Meta-Analysen kommt Aguinis zu folgendem Schluss: „Theoretically, it is likely that low power has led to the incorrect conclusion that there is no moderating effect in numerous research domains“ (Aguinis 2004: 83). Im Kontext des linearen Regressionsmodells liegen zahlreiche methodische Studien vor, die dieses Teststärkeproblem und seine Determinanten untersuchen (siehe die Zitationen in Aguinis 2004). Für nicht-lineare Regressionsmodelle existiert dagegen keine entsprechend umfassende Literatur. Dabei ist aufgrund des zu Beginn beschriebenen Problems davon auszugehen, dass das Auffinden von variablen-spezifischen Interaktionseffekten in solchen Modellen noch um ein Vielfaches schwieriger ist.

Um dieser Vermutung nachzugehen und die empirische Prüfbarkeit der MFS-Hypothesen im logistischen Regressionsmodell beurteilen zu können, werden Monte-Carlo-Simulationen durchgeführt. Mittels dieser lässt sich sowohl die Wahrscheinlichkeit eines α -Fehlers als auch die Teststärke für verschiedene mögliche Szenarien ermitteln. Die Konzeption und die Ergebnisse dieser Analysen werden im Folgenden beschrieben.

Generierung der wahren Werte für unterschiedliche Szenarien

Statistische Schätzfehler können selten allein auf Basis realer Daten untersucht werden, da sich die Schätzungen (und damit etwaige Schätzfehler) auf *unbekannte* Populationsparameter beziehen (Mooney 1997: 77f.). In Monte-Carlo-Simulationen wird davon ausgegangen, diese Populationsparameter wären bekannt, um dann über wiederholte Stichprobenziehungen und entsprechende Parameterschätzungen die Güte des Verfahrens beurteilen zu können.

Die Festlegung der Populationsparameter erfolgt nach theoretischen und substantiellen Erwägungen bzw. dem Erkenntnisinteresse. Ausgangspunkt ist das folgende logistische Regressionsmodell (für die Population und ohne Indizierung nach Personen):

$$\ln\left(\frac{\Pr(Y = 1 | \mathbf{x})}{\Pr(Y = 0 | \mathbf{x})}\right) = \beta_0 + \beta_N x_N + \beta_A x_A + \beta_{IE} x_N x_A + \beta_K x_K \quad (E1)$$

Dabei bezeichnet $Y = 1$ die Ausführung der von der Norm N verlangten Handlung y und $Y = 0$ das Unterlassen dieser Handlung. x_N entspricht einer bestimmten Ausprägung der Normvariablen X_N , x_A einer Ausprägung der Anreizvariablen X_A und x_K einer Ausprägung einer Kontrollvariablen X_K .

Um die Parameter dieser Regressionsgleichung substantiell besser interpretieren zu können, werden alle Prädiktoren auf Werte zwischen 0 und 1 skaliert, wobei größere Werte einer stärkeren Ausprägung der Norminternalisierung bzw. des Anreizes entsprechen. Festgelegt werden müssen dann folgende Populationsparameter:

- Der Interaktionseffekt zwischen Norm- und Anreizvariablen: β_{IE} ,
- der partielle Effekt der Normvariablen, wenn die Anreizvariable den Wert 0 annimmt: β_N ,
- der partielle Effekt der Anreizvariablen, wenn die Normvariable den Wert 0 annimmt: β_A ,
- der partielle Effekt der Kontrollvariablen: β_K
- sowie die Konstante: β_0 .

Zur Überprüfung der Teststärke muss davon ausgegangen werden, dass die MFS-Interaktionshypothese wahr ist. Die Populationsparameter werden entsprechend festgelegt. Untersucht wird dann, wie groß die Wahrscheinlichkeit ist, mit Hilfe des logistischen Regressionsmodells einen entsprechenden, statistisch signifikanten Interaktionseffekt in einer Stichprobe zu finden. Mit Hilfe der Monte-Carlo-Simulationen soll zudem analysiert werden, in Abhängigkeit von welchen Faktoren die Teststärke variiert. Dazu werden unterschiedliche (wenngleich MFS-konforme) Szenarien auf der Ebene der *Population* betrachtet. Es werden entsprechend mehrere künstliche Datenstrukturen erstellt. Diese stellen unterschiedliche Populationsszenarien nach, die in quantitativen Analysen von Umfragedaten zu Grunde liegen könnten. Im Folgenden wird beschrieben und begründet, in welcher Weise die Populationsparameter variiert wurden, um zu den verschiedenen Datenstrukturen zu gelangen:

- $\beta_0 \in \{-4, -2.5, -1, 0.5\}$: Der Parameter β_0 entspricht der logarithmierten Chance von Fällen, die auf allen unabhängigen Variablen den minimalen Wert (0) aufweisen. Über unterschiedliche Werte dieser Konstanten lassen sich Unterschiede in der *Schiefe der Verteilung* der binären abhängigen Variablen in der Population betrachten. Die Schiefe hängt aber auch von den anderen Eigenschaften der Population (Werten anderer Parameter und Verteilungen der Variablen) ab. Die vier für β_0 eingesetzten Werte entsprechen daher nicht einer bestimmten Prozentzahl von $Y = 1$, sondern Niveauunterschieden in den Prozentzahlen. Für die Datenstrukturen ohne Kontrollvariable (siehe unten) ergeben sich folgende Werte: Für $\beta_0 = -4$ liegt der Median bei 33% (min = 8%, max = 64%), für $\beta_0 = -2.5$ bei 57% (min = 26%, max = 85%), für $\beta_0 = -1$ bei 82% (min = 61%, max = 95%) und für $\beta_0 = 0.5$ bei 95% (min = 87%, max = 99%).
- $\beta_N \in \{2, 6\}$: Zumeist ergibt sich ein starker bis sehr starker partieller Effekt der Normvariablen, wenn die Anreizvariable den Wert 0 annimmt. Denn aufgrund der Kodierung der Variablen gibt dieser Effekt an, um welchen Faktor sich die Chance bei Abwesenheit des positiven Anreizes erhöht, wenn die Norminternalisierung maximal anstatt minimal ist.³ Eingesetzt in die Exponentialfunktion ergeben die Werte 2 und 6 multiplikative Erhöhungen der Chance um das 7.39-fache und 403.43-fache. Diese Werte entsprechen also einer starken und äußerst starken Erhöhung der Chance. Der Einfachheit halber bezeichne ich die beiden Ausprägungen des Parameters im Folgenden als „schwachen Normeffekt“ und „starken Normeffekt“.
- $\beta_A \in \{0.5, 1.5, 3\}$: Es wird immer von $\beta_A > 0$ ausgegangen, da ein positiver Anreiz die Handlung y auszuführen betrachtet wird. Verglichen mit dem partiellen Effekt der Normvariablen ist derjenige der Anreizvariablen deutlich geringer anzusetzen. Internalisierte Normen, die ein bestimmtes Verhalten vorschreiben, sind zumindest im Rahmen von Umfragen meist sehr hoch mit dem selbstberichteten Verhalten

³ Bei maximaler Norminternalisierung ist die Wahrscheinlichkeit, die Handlung auszuführen, sehr groß, bei minimaler Norminternalisierung und Abwesenheit des positiven Anreizes dagegen sehr gering. Daher ergibt sich zumeist ein starker konditionaler Effekt der Normvariablen auf das Verhältnis dieser Wahrscheinlichkeiten (d.h. auf die Chance, die Handlung auszuführen).

oder entsprechenden Intentionen assoziiert. Anreize sind verglichen dazu weiter entfernte Determinanten. Die empirischen Korrelationen und partiellen Effekte sind daher meistens deutlich geringer, obwohl sie teilweise durchaus stark sein können. Dem wurde die durch die Wahl der drei Werte für β_A Rechnung getragen. Eingesetzt in die Exponentialfunktion ergeben die Werte 0.5, 1.5 und 3 multiplikative Erhöhungen der Chance um das 1.65-fache, 4.48-fache und 20.09-fache. Diese werden im Folgenden als „schwache“, „mittlere“ und „starke“ Anreizeffekte bezeichnet.

- $\beta_{IE} \in \{-0.5, -1.5, -3\}$: Auf Basis des MFS lässt sich die erwartete Stärke des Interaktionseffektes in der Population theoriegeleitet bestimmen: Er sollte so stark sein, dass der partielle Effekt der Anreizvariablen gleich null ist, wenn die Normvariable maximal ausgeprägt ist. Da beide Variablen auf das Einheitsintervall kodiert werden, gilt dann: $\beta_{IE} = -\beta_A$. Da ein positiver Anreiz betrachtet wird, nimmt der Parameter β_{IE} entsprechend negative Werte an, die betragsmäßig jeweils dem Wert für β_A entsprechen. Diese beiden Parameter werden also in den Analysen der Teststärke nicht unabhängig voneinander variiert.
- $\beta_K \in \{0, -2\}$: Im Rahmen von Tests der MFS-Hypothesen kann es sinnvoll sein, auch noch für Drittvariablen zu kontrollieren, da es praktisch unmöglich ist, alle theoretisch möglichen Einflussgrößen empirisch umfassend zu messen. Der Einschluss von Kontrollvariablen kann die Ergebnisse von Hypothesentests zum einen durch Multikollinearität beeinflussen. Dies ist bekannt und soll hier nicht weiter untersucht werden. Zum anderen können Kontrollvariablen die „Ausgangswahrscheinlichkeit“ für $Y = 1$ beeinflussen. Um diese zweite Einflussmöglichkeit möglichst isoliert zu untersuchen, wird im Folgenden von Nullkorrelationen zwischen der Kontrollvariablen und den anderen beiden unabhängigen Variablen ausgegangen. Es werden zwei Ausprägungen des partiellen Effektes der Kontrollvariablen untersucht. Um Szenarien *ohne* Kontrollvariable zu betrachten, wird sein Koeffizient auf 0 gesetzt. Für Szenarien mit Kontrollvariable wird von einem Koeffizienten von -2 ausgegangen, was einer Verringerung der Chance um das 7.39-fache entspricht. Dieser relative hohe Wert wurde gewählt, da hier nur eine, in realen Analysen aber häufig mehrere Kontrollvariablen berücksichtigt werden.

Da β_A und β_{IE} direkt voneinander abhängen und nur gemeinsam variiert werden, ergeben sich über die Variation von β_0 , β_N , β_A (bzw. β_{IE}) und β_K entlang der aufgeführten Ausprägungen 48 verschiedene Populationen. Diese umfassen jeweils 150 000 Fälle (ebenso Cheng und Long 2007).

Des Weiteren müssen Annahmen über die gemeinsamen Verteilungen der unabhängigen Variablen getroffen werden. Vereinfachend wird angenommen, dass die Varianzen der unabhängigen Variablen unabhängig voneinander sind. Die Korrelation zwischen der Anreiz- und der Normvariablen wird auf einen realistischen Wert in Höhe von 0.2 festgelegt. Zudem wird für die Norm- und die Anreizvariablen von einer schiefen Verteilung ausgegangen, die den Verteilungen in den Stichproben der empirischen Anwendungen ähnelt. Dafür wurde eine Zufallsvariable auf Basis einer χ^2 -Verteilung mit 1.2 Freiheitsgraden gebildet und auf das Einheitsintervall rekodiert.⁴ Inhaltlich wird damit angenommen, dass die Mehrzahl der Fälle sowohl eine starke Normverankerung als auch einen starken positiven Anreiz zur Wahl des betrachteten Verhaltens aufweisen. Für die Kontrollvariable wurde der Einfachheit halber von einer Gleichverteilung ausgegangen.

Nachdem die unabhängigen Variablen erzeugt wurden und das Muster ihres Einflusses durch entsprechende Populationsparameter festgelegt wurde, können die Werte auf der abhängigen Variablen berechnet werden. Dies geschieht auf der Basis folgender Umformung von Gleichung E1:

$$\Pr(Y = 1 | x) = \frac{\exp(\beta_0 + \beta_N x_N + \beta_A x_A + \beta_{IE} x_N x_A + \beta_K x_K)}{1 + \exp(\beta_0 + \beta_N x_N + \beta_A x_A + \beta_{IE} x_N x_A + \beta_K x_K)} \quad (E2)$$

⁴ Die Simulationen zur Teststärke wurden auch für andere Verteilungsannahmen durchgeführt, die aber weniger realistisch sind. Folgen die Variablen einer *Normalverteilung*, so geht die Teststärke unter allen Bedingungen gegen 0. Der Grund ist, dass die Stichproben dann nur eine verschwindend geringe Anzahl von Fällen mit Extremwerten enthalten, welche für den Test der MFS-Hypothese entscheidend sind. Dies liegt daran, dass der Wertebereich einer Normalverteilung von $-\infty$ bis $+\infty$ reicht und sich die Verteilung nach den Seiten hin extrem ausdünn (im Gegensatz zu realistischen Verteilungen von Umfrage-Items). Höhere Teststärken ergeben sich, wenn man von *Gleichverteilungen* ausgeht, da in diesem Fall die Enden der Verteilungen relativ stark besetzt sind. Die Teststärke liegt teilweise deutlich höher als bei der χ^2 -Verteilung, die den im Folgenden berichteten Ergebnissen zu Grunde liegt. Auch gleich verteilte Variablen sind aber relativ unrealistisch, da sich – gerade bei der Normvariablen – zumeist schiefe Verteilungen ergeben.

Auf diese Weise wird hier für jeden Fall die wahre Wahrscheinlichkeit berechnet, Handlung y auszuführen ($Y = 1$). Man könnte auch von der wahren Handlungsdisposition sprechen. „Wahr“ deshalb, weil wir uns immer noch auf der Ebene der Population befinden. Dass es dennoch um Wahrscheinlichkeiten geht, reflektiert die Annahme, dass auch in der Population keine deterministischen, sondern probabilistische Beziehungen bestehen.

Im nächsten Schritt werden die Realisationen dieser Zufallsvariablen zugewiesen, also die Werte auf der binären abhängigen Variablen Y . Dafür wird eine gleichverteilte Zufallszahl z zwischen 0 und 1 generiert. Für jeden Fall in der Population wird Y auf 0 gesetzt, sofern $z \leq \Pr(Y = 1 \mid x)$ bzw. umgekehrt Y auf 1 gesetzt, falls $z > \Pr(Y = 1 \mid x)$ (ebenso Cheng und Long 2007: 590).

Design der Simulationen

Im Rahmen der Monte-Carlo-Simulation werden aus jedem der 48 verschiedenen Populationsdatensätze 1000 Stichproben gezogen (mit Zurücklegen). In jeder der gezogenen Stichproben wird das in Gleichung E1 dargestellte Regressionsmodell geschätzt und festgehalten, ob die MFS-Hypothese bestätigt wird oder nicht. Die MFS-Hypothese gilt als bestätigt (bzw. die Nullhypothese als zurückgewiesen), wenn drei Bedingungen erfüllt sind. Erstens muss der Interaktionseffekt statistisch signifikant sein. Als Kriterium kommen dabei der konventionelle p -Wert von 0.05 sowie der p -Wert von 0.10 zum Einsatz. Zweitens muss der bedingte Effekte der Anreizvariablen positiv ($\hat{\beta}_A > 0$) und drittens der Interaktionseffekt negativ ($\hat{\beta}_{IE} < 0$) geschätzt werden. Der Anteil der 1000 Stichproben, in denen die MFS-Hypothese in dieser Weise bestätigt werden kann, entspricht der Teststärke.

Um den Einfluss der Stichprobengröße abzuschätzen, wird diese Simulation für zwei unterschiedliche Stichprobengrößen durchgeführt: $N = 500$ und $N = 2000$. Diese Größenordnungen sind typisch für Datensätze in der Umfrageforschung, die angemessene Indikatoren enthalten, um MFS-Interaktionshypothesen testen zu können.

Hinsichtlich der Ergebnisse der Simulationen bestanden von Anfang an eine Reihe theoretischer Erwartungen. Zunächst existieren drei Faktoren, von denen bekannt ist, dass sie die Teststärke jedes inferenzstatistischen

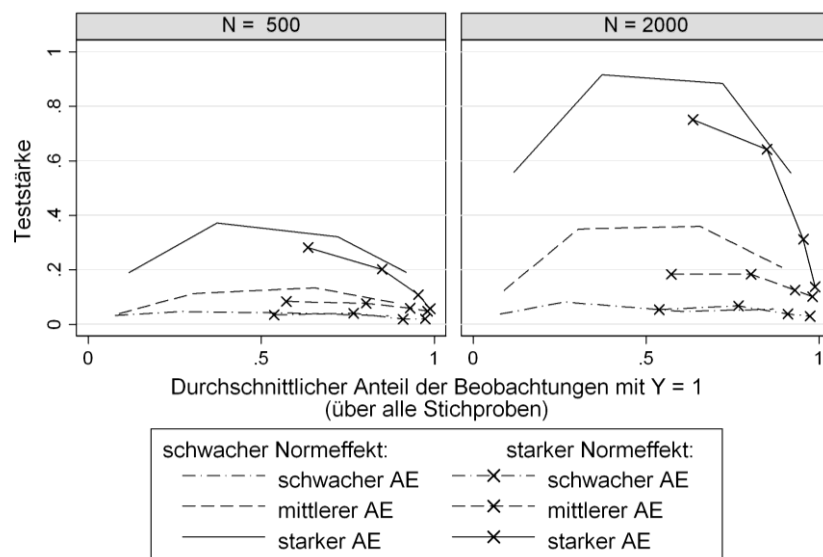
Tests beeinflussen. Danach ist die Teststärke umso geringer, je konservativer (also niedriger) der kritische α -Wert gewählt wird, je kleiner die Stichprobe und je schwächer der zu testende Effekt in der Population ist (Aguinis 2004: 69). Auf Basis der zu Beginn dargestellten Überlegungen lässt sich für unseren spezifischen Anwendungsfall noch eine weitere Erwartung formulieren: Es ist davon auszugehen, dass die Teststärke der MFS-Interaktionshypothese umso geringer ist, je stärker die Verteilung der „Ausgangswahrscheinlichkeiten“ $\Pr(Y = 1 \mid x)$ vom mittleren Wert von 0.5 abweicht.

Ergebnisse der Simulationen zur Teststärke

Abbildung 2 fasst erste Ergebnisse hinsichtlich der Teststärke und ihrer Determinanten zusammen. Dabei wird zunächst ein kritischer α -Wert von 0.05 verwendet und davon ausgegangen, dass kein Einfluss von Kontrollvariablen besteht. Interpretiert werden sollten nur die „Knicke“ der abgebildeten Linien, da nur diese Simulationsergebnisse darstellen. Die Linien wurden nur der Übersichtlichkeit halber durchgezogen.

Auf den ersten Blick fällt auf, dass die Teststärke mit wenigen Ausnahmen inakzeptabel gering ausfällt. Cohen (1992) empfiehlt, dass man eine Teststärke von mindestens 0.8 anstreben sollte. Das bedeutet, dass die Wahrscheinlichkeit, einen tatsächlich bestehenden Effekt auch zu bestätigen, bei 80 Prozent oder höher liegen sollte. Bei einer Stichprobengröße von $N = 500$ ist die Teststärke nicht einmal halb so groß und liegt in der Mehrzahl der Fälle sogar unter 0.2. Für $N = 2000$ ergeben sich wie erwartet höhere Teststärken, aber mit wenigen Ausnahmen fallen auch diese deutlich zu niedrig aus.

Abbildung 2: Teststärke in Abhängigkeit von der Verteilung der abhängigen Variablen in der Grundgesamtheit, der Stärke des Normeffektes und der Stärke des Anreizeffektes (AE) für $N = 500$ und $N = 2000$ bei Verwendung des 5%-Signifikanzkriteriums und ohne Kontrollvariable



Überwiegend ergibt sich somit ein ausgesprochen negativer Befund. Nichtsdestotrotz lässt sich eine ganze Bandbreite an Werten feststellen. Die Teststärke variiert von nahezu 0 bis zu akzeptablen Werten zwischen nahezu 80 und über 90 Prozent. Hinsichtlich der Determinanten der Teststärke lassen sich neben dem Effekt der Stichprobengröße einige weitere Ergebnisse festhalten:

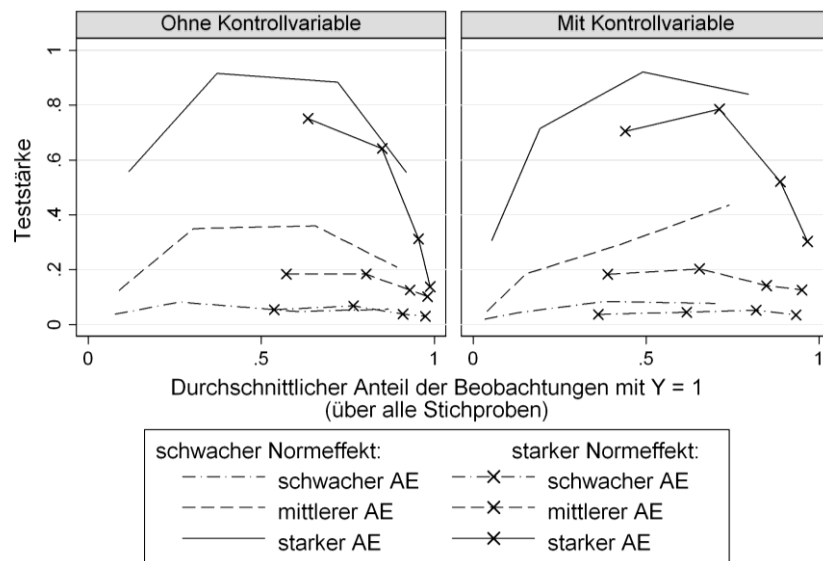
- In allen Szenarien, in denen akzeptable Werte erreicht werden, liegt ein starker Anreizeffekt in der Population vor. Ein aussagekräftiger Test der MFS-Hypothese ist demnach aussichtslos, sofern Anreizvariablen mit nur schwachem bis mittlerem Effekt betrachtet werden. Dieses Ergebnis steht im Einklang mit der Analyse von Rogers (2002). Dieser weist für lineare Regressionsmodelle mathematisch nach, dass Interak-

tionseffekte nur dann eine hohe Teststärke besitzen können, wenn der zu moderierende Effekt stark mit der abhängigen Variablen zusammenhängt.

- Es zeigt sich der erwartete, umgekehrt U-förmige Zusammenhang zwischen der mittleren Ausgangswahrscheinlichkeit $\Pr(Y = 1 \mid x)$ und der Teststärke. Liegt die durchschnittliche Ausgangswahrscheinlichkeit deutlich unterhalb oder oberhalb von 0.5, so führt dies zu einer merklich niedrigeren Teststärke. In Stichproben (oder Populationen), in denen die abhängige Variable extrem schief verteilt ist, sind die Bedingungen für eine Bestätigung der MFS-Hypothese demnach häufig nicht gegeben.
- Ein geringerer Einfluss geht von dem konditionalen Effekt der Normvariablen aus. Ein höherer konditionaler Effekt führt zwar insgesamt zu einer deutlich niedrigeren Teststärke. Hauptsächlich ist dieser Effekt aber durch die stärkere Schiefe der Verteilung der abhängigen Variablen vermittelt.

Ein aussagekräftiger Test der MFS-Hypothese erscheint demnach insbesondere dann möglich, wenn die dichotome abhängige Variable nicht zu schief verteilt ist, die Stichprobe annähernd 2000 Fälle umfasst und eine erklärungsstarke Anreizvariable vorliegt. Wie aber ist zu verfahren, wenn diese Bedingungen nicht gegeben sind? Eine methodisch saubere und nahe liegende Möglichkeit ist, von einem Test der Hypothese in solchen Situationen abzusehen. Sie ist jedoch unbefriedigend, wenn keine alternativen Daten verfügbar sind und eine Neuerhebung aus Kosten- oder anderen Gründen nicht unmittelbar möglich ist.

Abbildung 3: Teststärke in Abhängigkeit von der Relevanz einer Kontrollvariablen, der Verteilung der abhängigen Variablen in der Grundgesamtheit, der Stärke des Normeffektes und der Stärke des Anreizeffektes (AE) bei Verwendung des 5%-Signifikanzkriteriums und $N = 2000$



Eine Alternative besteht in der Identifikation und Einführung relevanter Kontrollvariablen. In Abbildung 3 sind die bekannten Ergebnisse (für $N = 2000$ ohne Kontrollvariable) den Ergebnissen für entsprechende Populationsdatensätze gegenübergestellt, in denen eine Kontrollvariable mit stark negativem Effekt angenommen wurde. Die Kontrollvariable bewirkt, dass die durchschnittliche Ausgangswahrscheinlichkeit für $Y = 1$ teilweise stark absinkt. Je nachdem, ob sich die durchschnittliche Ausgangswahrscheinlichkeit *ohne* Kontrollvariable unterhalb oder oberhalb von 0.5 befand, führt dies zu einer deutlichen Verringerung oder Erhöhung der Teststärke.

Als Beispiel für diesen Zusammenhang seien die Szenarien mit starkem konditionalem Anreizeffekt und schwachem konditionalem Normeffekt betrachtet (durchgezogene Linien ohne Kreuze). In der linken Graphik (ohne Kontrollvariable) ergibt sich für das Szenario mit dem höchsten Wert der Konstanten eine mittlere Ausgangswahrscheinlichkeit von 0.92 (siehe den

am weitesten rechts gelegenen Punkt auf dieser Linie). Die Teststärke liegt dadurch bei nur 0.56. Wie an der rechten Graphik ersichtlich, führt der Einschluss der Kontrollvariablen zu einer deutlichen Reduktion der mittleren Ausgangswahrscheinlichkeit (auf 0.80), wodurch die Teststärke auf den akzeptablen Wert von 0.84 ansteigt.

Vor dem beliebigen Einschluss von Kontrollvariablen muss allerdings gewarnt werden. Vor allem ist ein häufig wiederholtes Testen der MFS-Hypothese unter Variation der kontrollierten Drittvariablen zu vermeiden, da dies den tatsächlichen α -Fehler bekanntermaßen stark erhöhen kann (sog. „capitalizing on chance“).

Ergebnisse der Simulationen zum α -Fehler

Eine weitere Möglichkeit, eine höhere Teststärke herbeizuführen, besteht in der Anpassung des kritischen α -Niveaus. Das bislang zu Grunde gelegte Niveau fordert, dass die Wahrscheinlichkeit, fälschlicherweise vom Vorhandensein eines Interaktionseffekts auszugehen, kleiner als fünf Prozent ist. Um zu überprüfen, ob das Risiko eines α -Fehlers in unserem Anwendungsfall tatsächlich diesem Wert entspricht, wurden weitere Monte-Carlo-Simulationen durchgeführt. Diese beziehen sich nicht auf die Teststärke, sondern auf den α -Fehler. Ausgangspunkt war der Gedanke, dass das geringe Potential des logistischen Regressionsmodells, die vom MFS vorhergesagten variablenspezifischen Interaktionseffekte zu entdecken, zu einer Überschätzung des α -Fehlers führen könnte.

Die Simulationen folgten wiederum dem oben beschriebenen Design. Der einzige Unterschied bestand darin, dass der Koeffizient des Produktterms, β_{IE} , in allen 48 Populationsdatensätzen auf 0 gesetzt wurde.⁵ Das heißt, es wurde dieses Mal davon ausgegangen, dass die MFS-Hypothese falsch ist. Der Anteil der Stichproben, in denen die Nullhypothese fälschlicherweise dennoch verworfen wird, entspricht der Höhe des *tatsächlichen* α -

⁵ Es ergeben sich wiederum 48 Populationsdatensätze, da der Koeffizient des Produktterms bislang gemeinsam mit dem Koeffizienten der Anreizvariablen, β_A , variiert wurde. In den Monte-Carlo-Simulationen zum α -Fehler variiert nun allein der Koeffizient der Anreizvariablen.

Fehlers. Theoretisch sollte dieser Fehler bei Verwendung des 5%- bzw. 10%-Signifikanzniveaus ($p < 0.05$ bzw. $p < 0.10$) in fünf Prozent bzw. zehn Prozent aller Stichproben auftreten.

In Tabelle 1 sind die Ergebnisse zum α -Fehler denjenigen zur Teststärke gegenübergestellt. Im Unterschied zur Teststärke variiert die Höhe des α -Fehlers insgesamt sehr wenig zwischen den betrachteten Szenarien. Entgegen der erwähnten Vermutung entspricht der tatsächliche durchschnittliche α -Fehler in allen Szenarien den Signifikanzkriterien. Die sich ergebenden Abweichungen sind von vernachlässigbarer Größenordnung.

Auch die Monte-Carlo-Simulationen zur Teststärke wurden für das 10%-Signifikanzkriterium ausgewertet (siehe ebenfalls Tabelle 1). Wie zu erwarten, erhöht dieses weniger konservative Kriterium die Teststärke. Durchschnittlich ergibt sich eine sehr deutliche Erhöhung um durchschnittlich neun Prozentpunkte (maximal um 13 Prozentpunkte) verglichen mit der Teststärke bei Verwendung des 5%-Signifikanzkriteriums. In Prozentpunkten ist der Gewinn an Teststärke somit deutlich größer als die Erhöhung des Risikos eines α -Fehlers.

Dennoch mag der Ausweg, einen weniger konservativen Test zu akzeptieren, nicht ohne Weiteres gangbar erscheinen. Es ist allerdings zu bedenken, dass die Wahrscheinlichkeit, die MFS-Hypothese fälschlicherweise zu akzeptieren, tatsächlich nur halb so groß ist wie der bislang betrachtete α -Fehler. Dieser wurde für einen *zweiseitigen* Test berechnet: In der Hälfte der Fälle entspricht das fälschliche Verwerfen der Nullhypothese daher einem Interaktionseffekt, dessen Vorzeichen der MFS-Hypothese widerspricht. Bei einem 5%-Signifikanzniveau beträgt das Risiko, die MFS-Hypothese fälschlicherweise zu akzeptieren, demnach nur 2.5 Prozent, bei einem 10%-Signifikanzniveau fünf Prozent. Insofern das MFS gerichtete Hypothesen impliziert, erscheint es durchaus möglich, die Verwendung eines einseitigen Tests zu rechtfertigen (Long 1997: 70). Bei zweiseitigen Tests, wie sie in den folgenden empirischen Anwendungen nichtsdestotrotz berichtet werden, entspricht dies einem 10%-Signifikanzniveau (siehe zudem Jones und Tukey 2000).

Tabelle 1: Ergebnisse der Monte-Carlo-Simulationen: Tatsächlicher durchschnittlicher α -Fehler und Teststärke bei Verwendung des 5%- und 10%-Signifikanzkriteriums in Abhängigkeit von der Stichprobengröße, der Stärke des Normeffektes und der Verteilung der abhängigen Variablen in der Grundgesamtheit (Annahme: Starker Anreizeffekt)

			Kriterium: p < 0.05		Kriterium: p < 0.10	
N	Norm- effekt†	Pr(Y = 1) ††	∅	∅	∅	∅
			α-Fehler	Teststärke	α-Fehler	Teststärke
500	schwach	(,3, .7]	0.05	0.25	0.11	0.36
			n = 1	(.22, .28), n = 2	n = 1	(.33, .39), n = 2
			(.7,1)	0.04	0.15	0.09
	stark	(,3, .7]	(.02, .06), n = 7	(.06, .27), n = 6	(.04, .11), n = 7	(.11, .40), n = 6
			0.05	0.37	0.10	0.50
			(.04, .05), n = 3	(.37, .37), n = 2	(.10, .10), n = 3	(.50,.51), n = 2
2000	schwach	(,3, .7]	0.05	0.73	0.10	0.82
			n = 1	(.71, .75), n = 2	n = 1	(.80, .85), n = 2
			(.7,1)	0.05	0.45	0.09
	stark	(,3, .7]	(.04, .06), n = 7	(.14, .79), n = 6	(.08, .11), n = 7	(.24, .86), n = 6
			0.05	0.92	0.09	0.96
			(.04, .06), n = 3	(.92, .92), n = 2	(.08, .10), n = 3	(.96, .97), n = 2
			0.05	0.76	0.09	0.85
			(.04, .06), n = 4	(.56, .88), n = 3	(.08, .10), n = 4	(.69, .94), n = 3
Ungewichteter Durchschnitt			0.05	0.49	0.10	0.58

Anmerkungen: In Klammern: (Minimum, Maximum); n ist die Anzahl der Grundgesamtheiten in der Kategorie; [†]Effekt der Normvariablen in Grundgesamtheiten *ohne* Interaktionseffekt bzw. konditionaler Effekt der Normvariablen unter der Bedingung minimalen Anreizes in Grundgesamtheiten *mit* Interaktionseffekt. ^{††}Durchschnittliche vorhergesagte Wahrscheinlichkeit für Y = 1 in der Grundgesamtheit

Schlussfolgerungen

Aus den Resultaten der Monte-Carlo-Simulationen ergibt sich eine Reihe zentraler und untereinander eng zusammenhängender Konsequenzen für die folgenden Sekundäranalysen. Generell sollten die Standardfehler als Maß der statistischen Unsicherheit betrachtet werden. Auch wenn die Ergebnisse

von Signifikanztests berichtet werden, ist eine zu strikte Befolgung dieser Konventionen zu vermeiden. *Generell* fragwürdig ist vor allem die Praxis, die Nullhypothese (allein) auf Basis der p-Werte als verworfen oder bestätigt anzusehen (Cohen 1994; Gigerenzer, Krauss und Vitouch 2004; Jones und Tukey 2000; Krantz 1999; Loftus 1991; Nickerson 2000). Ausgehend von einem aufgeklärten Verständnis von Signifikanztests sollte stattdessen die Sicherheit der Schätzung im Gesamtzusammenhang der Ergebnisse betrachtet werden.

In logistischen (und anderen nicht-linearen) Regressionsmodellen sollte das Hauptaugenmerk zunächst auf dem Vorzeichenmuster und der relativen Stärke der betrachteten Effekte liegen. Wenn diese den Vorhersagen des MFS entsprechen, liegt ein erstes Anzeichen für die Gültigkeit der MFS-Hypothesen vor. Widersprechen die Koeffizienten dagegen den theoretischen Erwartungen, so ist dies ein ernst zu nehmendes Anzeichen dafür, dass die MFS-Hypothesen nicht zutreffen. Berücksichtigt werden sollten zudem die Eigenschaften der Stichprobe und der Population, welche die Teststärke prägen. Es spricht beispielsweise insgesamt deutlich für eine Forschungshypothese, wenn unter Bedingungen niedriger Teststärke theoriekonforme Koeffizienten geschätzt werden und diese p-Werte von 0.07 (bei zweiseitiger Testung) aufweisen.

Hinsichtlich des Tests der MFS-Hypothesen ist insbesondere bei Datensätzen mit extrem schief verteilter abhängiger Variable und relativ geringer Fallzahl ($N \approx 500$) mit einer äußerst oder zumindest relativ geringen Teststärke zu rechnen (nach Tabelle 1 zwischen 0.11 und 0.44 selbst bei einem 10%-Signifikanzniveau). Generell gilt, dass Tests der MFS-Hypothese nur aussagekräftig sind (d.h. potentiell über eine gewisse Teststärke verfügen), wenn mindestens eine *erklärungsstarke* Anreizvariable identifiziert werden kann. Von Fragen der statistischen Signifikanz abgesehen, kann es freilich dennoch von Interesse sein, auch für weniger erklärungsstarke Anreizvariablen Interaktionseffekte zu spezifizieren und die sich ergebenden konditionalen Effekte zu betrachten.

Bei der Interpretation der Ergebnisse der folgenden Sekundäranalysen sollten diese Prinzipien zu Grunde gelegt werden. Auch bei der Erhebung von Primärdaten zum Test von MFS-Hypothesen können die hier durchgeführten Teststärkeanalysen nützlich sein. So können im Rahmen der Datenerhebung verschiedene Vorkehrungen für eine hinreichende Teststärke getroffen werden. Dazu zählen etwa die vorherige Identifikation hinreichend

erklärungsstarker Anreizvariablen und die Optimierung ihrer Messung im Rahmen von Pre-Tests. Über spezielle Sampling-Strategien könnte man überdies sicherstellen, dass die abhängige Variable relativ ausgewogen verteilt ist und dass Fälle mit den theoretisch besonders relevanten, extremen Merkmalskombinationen in ausreichender Zahl verfügbar sind (McClelland und Judd 1993). Allerdings sind derartige Maßnahmen gegen den Nachteil mangelnder Repräsentativität abzuwägen (siehe Aguinis 2004: 89).

Literatur

- Achen, Christopher H., 2002: Toward a New Political Methodology: Microfoundations and ART. *Annual Review of Political Science* 5: 423-450.
- Aguinis, Herman, 2004: *Regression Analysis for Categorical Moderators*. New York: The Guilford Press.
- Ai, Chunrong und Edward C. Norton, 2003: Interaction Terms in Logit and Probit Models. *Economics Letters* 80: 123-129.
- Allison, Paul D., 1999: Comparing Logit and Probit Coefficients Across Groups. *Sociological Methods & Research* 28: 186-208.
- Braumoeller, Bear F., 2003: Causal Complexity and the Study of Politics. *Political Analysis* 11: 209-233.
- , 2004: Boolean Logit and Probit in Stata. *The Stata Journal* 4: 436-441.
- Brüderl, Josef, 2004: Die Überprüfung von Rational-Choice-Modellen mit Umfragedaten. S. 163-180 in: Andreas Diekmann und Thomas Voss (Hg.), *Rational-Choice-Theorie in den Sozialwissenschaften. Anwendungen und Probleme*. München: Oldenbourg.
- Cheng, Simon und J. Scott Long, 2007: Testing for IIA in the Multinomial Logit Model. *Sociological Methods & Research* 35: 583-600.
- Cohen, Jacob, 1992: A power primer. *Psychological Bulletin* 112: 155-159.
- , 1994: The earth is round ($p < 0.05$). *American Psychologist* 49: 997-1003.
- Gigerenzer, Gerd, Stefan Krauss und Oliver Vitouch, 2004: The Null Ritual. What You Always Wanted to Know About Significance Testing but Were Afraid to Ask. S. 391-408 in: David Kaplan (Hg.), *The Sage Handbook of Quantitative Methodology for the Social Sciences*. Thousand Oaks, CA: Sage.
- Jones, Lyle V. und John W. Tukey, 2000: A Sensible Formulation of the Significance Test. *Psychological Methods* 5: 411-414.
- Kohler, Ulrich und Frauke Kreuter, 2008: *Datenanalyse mit Stata. Allgemeine Konzepte der Datenanalyse und ihre praktische Anwendung*. München: Oldenbourg.

-
- Krantz, David H., 1999: The Null Hypothesis Testing Controversy in Psychology. *Journal of the American Statistical Association* 94: 1372-1381.
- Loftus, Geoffrey R., 1991: On the Tyranny of Hypothesis Testing in the Social Sciences. *Contemporary Psychology* 36: 102-105.
- Long, J. Scott, 1997: Regression Models for Categorical and Limited Dependent Variables. Thousand Oaks, CA: Sage.
- Maier, Gunther und Peter Weiss, 1990: Modelle diskreter Entscheidungen. Theorie und Anwendungen in den Sozialwissenschaften. Wien: Springer.
- Manski, Charles F., 1995: *Identification Problems in the Social Sciences*. Cambridge: Harvard University Press.
- McClelland, Gary H und Charles M. Judd, 1993: Quantitative Methods in Psychology: Statistical Difficulties of Detecting Interactions and Moderator Effects. *Psychological Bulletin* 114: 376-390.
- Mood, Carina, 2010: Logistic Regression: Why We Cannot Do What We Think We Can Do, and What We Can Do About It. *European Sociological Review* 26: 67-82.
- Mooney, Christopher Z., 1997: *Monte Carlo Simulation*. Thousand Oaks, CA: Sage.
- Nagler, Jonathan, 1994: Scobit: An Alternative Estimator to Logit and Probit. *American Journal of Political Science* 38: 230-255.
- Nickerson, Raymond S., 2000: Null Hypothesis Significance Testing: A Review of an Old and Continuing Controversy. *Psychological Methods* 5: 241-301.
- Rogers, William M., 2002: Theoretical and mathematical constraints of interaction regression models. *Organizational Research Methods* 5: 212-230.

Die Erklärung sozialen Handelns
Grundlagen und Anwendung einer integrativen Theorie
Kroneberg, C.
2011, X, 348 S., Softcover
ISBN: 978-3-531-17389-4