

Thomas Bräuninger

Fruchtbare Boden oder doch nur Wüste?

Eine Replik auf Tina Freyburg

Im Beitrag »Demokratisierung durch Kooperation? Funktionale Kooperation mit autoritären Regimen und Sozialisation in demokratischem Regieren« argumentiert Tina Freyburg, die funktionale Kooperation demokratischer mit autoritären Staaten könne subtile Demokratisierungsprozesse in letzteren auslösen. Ein Beispiel hierfür sei die demokratische Sozialisation von Regierungsbeamten und -beamtinnen in autoritären Regimen, die im Rahmen des Twinning-Programms der Europäischen Union mit Prinzipien des demokratischen Regierens vertraut gemacht werden. Die Auswertung von Daten einer Befragung von marokkanischen Staatsbeamten scheint die These eines demokratischen Wandels autoritärer Regime durch Kooperation zu stützen. In der vorliegenden Replik bespreche ich einige der methodischen Probleme, die sich bei einem Nachweis der Änderung von grundlegenden Einstellungen aus Befragungs- und quasi-experimentellen Daten ergeben sowie die Implikationen daraus für Freyburgs Beitrag.

1. Demokratischer Wandel durch Kooperation

In ihrem ZIB-Beitrag »Demokratisierung durch Kooperation? Funktionale Kooperation mit autoritären Regimen und Sozialisation in demokratischem Regieren« untersucht Tina Freyburg (2011), ob eine Form der funktionalen Kooperation von demokratischen und autoritären Staaten auf der Verwaltungsebene dazu führen kann, dass Staatsbeamte in autoritären Staaten eine demokratische Sozialisation erfahren und positive Einstellungen zu demokratischen Formen des Regierens ausbilden. Gegenstand der empirischen Untersuchung sind marokkanische Staatsbeamte, die in Kooperationsprojekten im Rahmen des Twinning-Programms der Europäischen Union (EU) teilgenommen haben, mit dem die EU die Modernisierung von Verwaltungsstrukturen in assoziierten und Mittelmeeranrainerstaaten fördern will. Der Beitrag verbindet in geradezu idealer Weise politikwissenschaftliche und politisch-praktische Relevanz. Aus politikwissenschaftlicher Sicht steht sicherlich die Frage nach den Bedingungen für eine demokratische Sozialisation im Vordergrund. Können Individuen, die ihre primäre und politische Sozialisation in autoritären Regimen erfahren haben, Demokratie erlernen und wenn ja, gibt es Bedingungen hierfür? Hinzu kommt die methodische Frage, wie durch das Instrument der Befragung mögliche Einstellungsänderungen festgestellt werden können. Wie kann eine tat-

sächliche Änderung von Einstellung von einfachem *cheap talk* und sozial erwünschtem Antwortverhalten unterschieden werden?¹

Zum anderen hat die Frage eine praktische Relevanz. Das *Twinning*-Programm der EU war 2004 auf die Mittelmeieranrainerstaaten erweitert worden, nachdem das Programm offenbar in den osteuropäischen Beitrittsländern große Erfolge erzielt hatte. Man wüsste nur zu gerne, ob und wie sich ein Erfolg des Programms manifestiert bzw. die aufgewendeten Mittel sinnvoll eingesetzt wurden. Hier geht es zunächst um das vorrangige – etwas eigennützige – Ziel der EU, Beitritts- und Anrainerstaaten auf die Spielregeln des *aquis communautaire* einzuschwören, was wiederum Voraussetzung für den Beitritt bzw. eine vertiefte Kooperation mit der Europäischen Union ist. Weitergehend wird mit den *Twining*-Projekten aber auch die Hoffnung verbunden, die funktionale Kooperation mache die Verwaltungen in autoritären Staaten mit prinzipiellen Mustern und Regeln der demokratischen Entscheidungsfindung vertraut, was im besten Fall »subtile Demokratisierungsprozesse« (Freyburg 2011: 7) freisetze, wenn demokratische Regeln und Praktiken zunehmend im bürokratischen Alltag umgesetzt und in einem weiteren Schritt von gesellschaftlichen Akteuren dann auch so eingefordert würden. Darüber hinaus könnte man sogar die Frage stellen, ob und wenn ja, welchen entscheidenden Beitrag die Europäische Nachbarschaftshilfe zu der so hoffnungsvoll beginnenden Demokratisierung nach dem Arabischen Frühling 2011 in Nordafrika hat leisten können.

Ich werde im Folgenden argumentieren, dass die empirischen Daten, die der Analyse von Tina Freyburg zugrunde liegen, die Haupthypothesen des Beitrags nicht stützen: Es finden sich meines Erachtens weder Hinweise darauf, dass die Teilnahme an Kooperationsnetzwerken eine demokratische Sozialisation hervorruft, noch dass ein vorgängiger Kontakt mit demokratischen Prinzipien notwendig ist, um über die Anknüpfung an »mentale Wissensstrukturen« (Freyburg 2011: 26) eine nachfolgende Sozialisation zu erlauben. Eine Reanalyse der Daten legt hingegen nahe, dass es sich bei der Analyse zugrunde liegenden Population um eine selektive Auswahl von Beamten handelt, so dass ein Sozialisationsmechanismus der beschriebenen Art nicht geschlussfolgert werden kann. Und selbst wenn, den postulierten »primacy« Effekt (Searing 1976) stellt es in keinem Falle dar. Weiterhin spezifiziert die Mehrzahl der von der Autorin durchgeführten Regressionsanalysen falsche Kontrollgruppen, so dass die Interpretation (möglicherweise) durch Fehler von Typ I als auch Typ II gekennzeichnet ist, das heißt Fehlschlüsse, bei denen die Nullhypothese entweder fälschlicherweise nicht zurückgewiesen oder fälschlicherweise zurückgewiesen wird.

Ich will diese drei Punkte auf den kommenden Seiten ausführen. Dazu gehe ich wie folgt vor. Erstens diskutiere ich das theoretische Kernargument des Beitrags, ob und wann ein Effekt der funktionalen Kooperation auf die Einstellung der involvierten Beamten zu erwarten ist (Abschnitt 2). Anschließend fasse ich das empiri-

1 Ich gehe von einem einfachen, bipolar-kontinuierlichen Einstellungs begriff aus, der Einstellungen definiert als »psychological tendency that is expressed by evaluating a particular entity with some degree of favor or disfavor« (Eagly/Chaiken 1993: 1) und ambivalente Einstellungen nicht berücksichtigt. Dies entspricht Freyburgs Verwendung des Begriffes.

sche Design der Studie zusammen und bespreche dabei Probleme des quasi-experimentellen Designs (Abschnitt 3). Drittens argumentiere ich, dass meines Erachtens und im Widerspruch zur Interpretation der Autorin die im Artikel präsentierten zentralen Ergebnisse der empirischen Analysen die zwei zentralen Hypothesen des Beitrags nicht stützen und diskutiere mögliche Gründe für diesen Befund (Abschnitt 4).

2. Das Argument

Der Beitrag von Tina Freyburg ergänzt die bisherige Forschung zur internationalen Sozialisation, indem er individuelle Einstellungsänderungen von Beamten zum administrativ-demokratischen Regieren untersucht, die durch funktionale Kooperationen hervorgerufen werden können. Die Autorin gibt zu bedenken, dass es sich dabei – im Vergleich zu bisherigen Arbeiten – zwar um einen engen Gegenstandsbereich (Normen des administrativ-demokratischen Regierens) und eine kurze Kausalkette (Kooperation führt zu Einstellungswandel) handelt. Man kann ihr aber beipflichten, dass angesichts der vielen nur moderat positiven empirischen Befunde in der internationalen Sozialisationsliteratur, eine Hinwendung zu den Mikromechanismen solcher Prozesse dringend notwendig ist.

Drei Faktoren sollen im konkreten Untersuchungsfall den Nachweis solcher Einstellungsänderungen erleichtern. Erstens geht es nur um Normen, die Teil der professionellen Praktiken der Beamten sind. Zweitens werden diese Normen nicht explizit gefördert, sondern durch sachlich-professionelle Zusammenarbeit vorgelebt. Beide Faktoren machen das Auftreten von sozial erwünschtem Antwortverhalten in der nachfolgenden Befragungssituation weniger wahrscheinlich, wenngleich es freilich nicht ausgeschlossen werden kann. Drittens geht es um Einstellungswandel und nicht um weiterreichende Verhaltensänderungen. Kurzum, wenn es so etwas wie Sozialisationseffekte gibt, dann sollten sie hier zu beobachten sein.

Ausgangspunkt der Argumentation ist die Annahme, dass die soziale Interaktion von Beamten auf Arbeitsebene bedeutend ist für Prozesse der demokratischen Sozialisation. Die erste, daraus unmittelbar abgeleitete Hypothese eins lässt dann erwarten, dass »Staatsbeamte, welche in transgouvernementale Politiknetzwerke involviert waren, [...] eine positivere Einstellung zu demokratischem Regieren [haben] als ihre Kollegen, die nicht an diesen Netzwerken teilnahmen« (Freyburg 2011: 12).² Während hier also Niveauunterschiede zwischen Beamten mit und ohne Teilnahme an Kooperationsprojekten zu erwarten sind, stellt eine weitere Hypothese auf den Einfluss von anderen transnationalen Faktoren ab.

So ermöglichten Austauschprogramme (von Studierenden und Berufstätigen) sowie die Nutzung ausländischer Medien, »dass Bürger nichtdemokratischer Staaten

2 Eine andere Frage ist freilich, ob gleichzeitig auch Erfahrungstransfers in umgekehrter Richtung und damit Sozialisationsprozesse bei den europäischen Kooperationspartnern zu erwarten sind.

Entscheidungsfindungsprozesse in demokratischen Staaten direkt erfahren können« bzw. über »demokratische Praktiken« informiert werden. Staatsbeamte und Staatsbeamtinnen könnten so »bereits vor ihrer Teilnahme an transgouvernementalen Netzwerken Erfahrungen mit demokratischem Regieren gemacht haben« (Freyburg 2011: 12-13). Dieser »primacy effect« (Hooghe 2005: 866) [sei] entscheidend für Sozialisation«, so dass »Staatsbeamte, die bereits Erfahrungen mit demokratischem Regieren gemacht haben, eher dazu geneigt sind, ihre Einstellungen zu ändern, wenn sie erneut in Kontakt mit demokratischem Regieren kommen«. Aus diesem Argument wird eine zweite Hypothese abgeleitet, und zwar sei es

»wahrscheinlicher, dass die Teilnahme an transgouvernementalen Netzwerken einen positiven Einfluss auf die Einstellungen zu demokratischem Regieren von Staatsbeamten ausübt, wenn diese längere Zeit in einer westlichen Demokratie zu Ausbildungs- oder beruflichen Zwecken verbracht haben [Hypothese 2a] oder regelmäßig westliche Medien zur politischen Information nutzen [Hypothese 2b]« (Freyburg 2011: 13).

Mit anderen Worten lässt die zweite Hypothese erwarten, dass für Beamte mit Auslandserfahrung (bzw. Mediennutzung) der Netzwerkeffekt auf demokratische Einstellungen positiv und größer als für Beamten ohne Auslandserfahrung (bzw. Mediennutzung) ist.³

Diese Erwartung lässt sich jedoch nur schwer aus der Annahme eines *primacy effects* oder Primäreffekts herleiten. Nach Liesbet Hooghe (2005: 866) ist darunter Folgendes zu verstehen: »Socialization varies inversely with a person's age and experience. This is the primacy effect. Psychologists theorize that new experiences stick best when a person has few relevant prior experiences«. Neue Erfahrungen werden demnach im Lichte von bisherigen Erfahrungen interpretiert und natürlich sind das (biologische) Alter des Individuums sowie die Menge seiner bisherigen Erfahrungen ein guter *proxy* für die Verfestigung von Einstellungen, die dann eben zu geringerem Maße geändert werden können. Wenn also, wie im Beitrag Freyburgs, die zweite Erfahrung demokratischen Regierens (in den transgouvernementalen Netzwerken) auf eine erste Erfahrung mit demokratischem Regieren (durch Auslandsaufenthalte oder Mediennutzung) aufbaut – oder an »mentale Wissensstrukturen anknüpft« (Freyburg 2011: 26) –, dann sollte der Effekt des Sozialisationsversuches am stärksten bei denjenigen sein, die erstmals in Kontakt mit den Prinzipien des demokratischen Regierens kommen (denn »new experiences stick best when a person has few relevant prior experiences« (Hooghe 2005: 866)). Das gilt für Beamten und Beamte, die erstmals an Kooperationsnetzwerken teilnehmen und keine Auslandserfahrung haben und auch für Individuen, welche eine erste Auslandserfahrung machen oder erstmals ausländische Medien nutzen. Da wir über keine Längsschnittdaten verfügen, aus denen sich eine Veränderung der Einstellungen eines einzelnen Individuums vor und nach der Kooperationserfahrung ablesen lassen, vergleichen wir zwei Individuen mit bzw. ohne Kooperationserfahrung. Für einen Kausalschluss muss dann allerdings die Annahme der »unit homo-

3 Hypothese 2 ist im Gegensatz zur ersten Hypothese probabilistisch formuliert, ohne dass dies aus theoretischen oder messtheoretischen Gründen notwendig erscheint. Ich interpretiere die Hypothese deshalb deterministisch.

geneity» (Holland 1986: 948) getroffen werden: eine konsistente Schätzung des Effekts gelingt nur, wenn die beiden Einheiten in den sonstigen Bedingungen vergleichbar sind. Kurzum, wenn y die Einstellung zum demokratischen Regieren bezeichnet, wobei größere Werte für positive Einstellungen stehen, dann ist zu erwarten, dass

$$y(Kooperation = 1 | Ausland = 0) - y(Kooperation = 0 | Ausland = 0)$$

$$> y(Kooperation = 1 | Ausland = 1) - y(Kooperation = 0 | Ausland = 1)$$

gilt. Nach den empirischen Befunden Freyburgs (Freyburg 2011: 25, Tabelle 3, Modell 3a) ist der erste Term aber signifikant kleiner als der zweite (-0,092 gegenüber 0,364 - 0,092=0,272). Das passt nun zu der in Hypothese zwei formulierten Erwartung. Es entspricht aber nicht der eines *primacy effects*, sondern deutet eher auf einen umgekehrten Mechanismus hin: »new experiences stick best when a person has many relevant prior experiences«. Ich werde im Folgenden argumentieren, dass sich auch für Hypothese zwei keine klare Evidenz in den empirischen Daten findet. In jedem Falle scheint das Argument, mit dem die Hypothese abgeleitet wird, fehlerhaft.

3. Untersuchungsdesign

Die Stärke und das Potenzial der Studie Freyburgs liegen darin, dass sie dem Phänomen der demokratischen Sozialisation auf der Ebene der Individuen, hier mit dem Instrument der Befragung, nachzuspüren versucht. Da Sozialisation ein Prozess ist, läge es nahe, die Einstellungsänderung eines Individuums zwischen zwei Zeitpunkten zu untersuchen, idealerweise in einem experimentellen Design im Vergleich einer Treatment- und einer Kontrollgruppe von Individuen. Kern des experimentellen Designs ist die randomisierte Zuweisung der Probanden zu Treatment- und Kontrollgruppen. Nur dann kann davon ausgegangen werden, dass sich die Individuen in beiden Gruppen im Durchschnitt und bei hinreichend großer Fallzahl nicht wesentlich hinsichtlich dritter, möglicherweise konfundierender Eigenschaften unterscheiden. Ist diese Bedingung erfüllt, dann kann der Unterschied in der abhängigen Variablen zuverlässig mit dem *treatment*-Effekt in Verbindung gebracht werden.

Ein experimentelles Design ist in der natürlichen Umgebung des marokkanischen Staatsapparates kaum zu realisieren, eine wiederholte Befragung, vor und nach der Teilnahme an einem *Twinning*-Projekt, ebenfalls schlecht. Tina Freyburg (2011: 17) wählt deshalb klug ein quasi-experimentelles Design, bei dem die Zuweisung zu den Gruppen nicht randomisiert erfolgt, aber mit der Vollerhebung von Beamtinnen in vorselektierten Abteilungen mit und ohne *Twinning*-Projekt-Erfahrung eine geeignete Kontrollgruppe generiert werden soll, »welche der Fokusgruppe in allen relevanten Charakteristika genügend ähnelt, außer, dass diese Gruppe nicht an Aktivitäten eines *Twinning*-Projektes teilgenommen hat« (Freyburg 2011: 17).

So sorgfältig dieses Design gewählt und implementiert ist, verbleiben zumindest zwei Probleme. Zunächst zeigt sich, dass die erhaltenen Fokus- und Kontrollgruppen nicht in allen als relevant erachteten Charakteristika ähnlich sind. Weibliche Staatsbeamte sind in der Fokusgruppe signifikant weniger häufig vertreten als in der Kontrollgruppe (mit Anteilen weiblicher Beamter in Fokus- und Kontrollgruppe von 25,9% bzw. 46,2%; $p=0,01$ des exakten Fisher-Tests), Nutzer ausländischer Medien sind stärker in der Fokusgruppe vertreten (86,4% bzw. 70,0%; $p=0,02$ des exakten Fisher-Tests). Man könnte nun beispielsweise vermuten, dass in patriarchalischen Gesellschaften es weibliche Bewerber generell schwerer haben, in den Staatsdienst zu kommen und dass dies vor allem denjenigen gelingt, die besonders regimetreu sind. Existiert ein solcher Selektionsbias, dann kann die Kombination von Geschlechterunterschieden im Ausgangsniveau und bei der Zuweisung zur *treatment*-Gruppe einen tatsächlichen Sozialisationseffekt neutralisieren oder aber verstärken mit der Folge, dass Hypothese eins fälschlicherweise zurückgewiesen oder bestätigt wird.

Selbst wenn kein Selektionsbias bei der Auswahl von Individuen in den Beamtenapparat vorliegt, kann es eine selektive, d.h. durch Drittvariablen verursachte Auswahl von Beamten und Beamtinnen zu Fokus- und Kontrollgruppe geben. Im vorliegenden Fall wird dies deshalb wahrscheinlich, weil zwei *treatments* (»Twining-Projektteilnahme« und »Auslandsaufenthalt«) eingeführt werden. Hinsichtlich der Teilnahme will die Autorin davon ausgehen, dass diese »auf der Grundlage objektiver Kriterien entschieden [wird] wie dem Zuständigkeitsbereich in der jeweiligen Abteilung und der professionellen Arbeitsleistung, nicht aufgrund von persönlichen Kontakten und Loyalitäten« (Freyburg 2011: 12). Freilich geht es auch hier nicht darum, ob persönliche Kontakte und Loyalitäten eine Rolle spielen, sondern darum, ob es *überhaupt* Faktoren für die Auswahl gibt, welche in die Analyse nicht eingehen. Hinsichtlich des Auslandsaufenthaltes ist die Annahme der Abwesenheit solcher Faktoren ungleich schwerer zu verteidigen. Da der Auslandsaufenthalt vor dem Eintritt die Ministerialbürokratie erfolgte (Freyburg 2011: 25), aber alle in der Analyse berücksichtigten Variablen sich auf Individualeigenschaften nach dem Eintritt beziehen, fußt jede Analyse auf der impliziten Annahme, das *treatment* Auslandsaufenthalt sei nach dem Zufallsprinzip auf Experimental- und Kontrollgruppen verteilt. Hat der sozio-ökonomische Status oder das Bildungsniveaus von späteren Beamten und Beamtinnen tatsächlich keinen Einfluss auf die Wahrscheinlichkeit eines Auslandsaufenthaltes? Das erscheint unwahrscheinlich, wir wissen es aber nicht.

Denn selbst wenn alle konfundierenden Variablen bekannt sind, ist ihre Kontrolle bei ungleicher Verteilung der Individuen über Fokus- und Kontrollgruppe nicht trivial. Die übliche, auch hier verwendete, Methode der Regressionsanalyse kann dies nur bedingt lösen (vgl. Gelman/Hill 2007: 200): Man nehme beispielsweise an, es soll der Effekt θ eines Twining-Projekts auf die Einstellung zum demokratischen Regieren y bestimmt werden, wobei die Dauer der Sozialisation im autoritären Staatsapparat x und die ausländische Mediennutzung z konfundierende Variablen

darstellen, die einen nicht-additiven Effekt auf y haben. Das entsprechende Regressionsmodell lautet dann:

$$y_i = b_0 + b_1 x_i + b_2 z_i + b_3 x_i z_i + \theta I_i + \varepsilon_i$$

wobei I_i die Indikatorvariable für die Teilnahme von Individuum i an der *treatment*-Gruppe ist. Der geschätzte *Treatment-Effekt* ergibt sich dann zu:

$$\hat{\theta} = \bar{y_T} - \bar{y_K} - b_1(\bar{x_T} - \bar{x_K}) - b_2(\bar{z_T} - \bar{z_K}) - b_3(\bar{x_T z_T} - \bar{x_K z_K}),$$

wobei $\bar{x_T}$ und $\bar{x_K}$ die Mittelwerte der jeweiligen Variablen für *treatment*- und Kontrollgruppen bezeichnen. Ignoriert man die konfundierende Variable vollständig, dann ist der *treatment-Effekt* um den Wert $b_1(\bar{x_T} - \bar{x_K}) - b_2(\bar{z_T} - \bar{z_K}) - b_3(\bar{x_T z_T} - \bar{x_K z_K})$ verzerrt; berücksichtigt man aus Unkenntnis des nicht-additiven Zusammenhangs x und z nur als additive Prädiktoren, dann ergibt sich immer noch eine Verzerrung von $b_3(\bar{x_T z_T} - \bar{x_K z_K})$. Mit anderen Worten, zur Bestimmung des *treatment-Effekts* in einem quasi-experimentellen Design gehört nicht nur Kenntnis der konfundierenden Variablen, sondern auch die Kenntnis ihres funktionalen Zusammenhangs mit der abhängigen Variablen.⁴

4. Empirische Analyse

Zur empirischen Untersuchung des Zusammenhangs von *Twining*-Projekten und demokratischem Einstellungswandel entwickelt Freyburg ein dreidimensionales theoretisches Konstrukt des demokratischen Regierens, das durch eine Batterie von acht Indikatoren operationalisiert wird, welche den Befragten dann in Form von fünfstufigen Likertskalen vorgelegt werden. Eine sorgfältig ausgeführte explorative Faktorenanalyse bestätigt die erwartete Struktur des Indikatorenraums und generiert über die Faktorwerte die drei Skalen der Partizipation, Verantwortlichkeit und Transparenz. Die Skala des übergeordneten Konstrukturts des demokratischen Regierens ergibt sich dann aus dem Mittelwert der drei Dimensionsskalen.⁵

Tabelle 3 des Freyburg-Artikels (2011: 25) berichtet dann den Kern der empirischen Analyse und ist Ausgangspunkt für die weiteren Schlussfolgerungen. Die Tabelle präsentiert die Ergebnisse von vier Regressionsmodellen, mit denen der Effekt von Kern- und Kontrollvariablen auf das übergeordnete Konzept des demokratischen Regierens untersucht wird. Die wesentlichen Ergebnisse und Schlussfolgerungen der Autorin sind: 1) Betrachtet man allein die Kooperation als

4 Matching-Verfahren bieten eine partielle Lösung dieses Problems (Gelman/Hill 2007: 206).

5 Der Indikator »ein Beamter sollte aktuelle Informationen zur Regierungspolitik zur Verfügung stellen« wird a priori der Partizipationsdimension (anstelle der Transparenzdimension) und umgekehrt der Indikator »ein Beamter sollte Bürgern die Möglichkeit bieten, ihre Ansichten als Beitrag zu Regierungsentscheiden vorzubringen« der Transparenzdimension (anstelle der Partizipationsdimension) zugeordnet. Dies passt dann gut zu den Ergebnissen der Faktorenanalyse (Anhang II), steht aber nicht im Einklang mit der theoretischen Herleitung, siehe auch Freyburg (2009: 10).

Prädiktorvariable (Modell 1), dann ergibt sich kein signifikanter Effekt, so dass Hypothese eins zurückgewiesen werden muss. 2) Bei der Berücksichtigung von Kontrollvariablen (Modell 2) ergibt sich ein positiver und signifikanter Koeffizient für die Teilnahme der Beamten »an anderen Programmen«, was »jedoch nicht notwendigerweise das Ergebnis der Teilnahme, sondern [...] auch durch die Häufung dieser Programme in weniger politisierten Bereichen hervorgerufen worden sein« (Freyburg 2011: 24) könne. 3) Berücksichtigt man weiterhin die Interaktion der Kernvariablen, dann ergibt sich laut Freyburg folgendes Bild: »Während für die Nutzung ausländischer Medien kein signifikanter bedingter Effekt von *Twinning*-Aktivitäten beobachtet werden kann, ist der Koeffizient des Interaktionsterms von Kooperation und Auslandsaufenthalt in Modell 3a positiv und statistisch signifikant, wie auch ein Wald-Test bestätigt ($p = .000; df=3; \chi^2=4793.418$). Dieser Befund stützt Hypothese 2 a, aber nicht Hypothese 2 b« (Freyburg 2011: 25).

Die Autorin schließt daraus, dass um »das Verständnis der Staatsbeamten von angemessenem Regieren signifikant beeinflussen zu können, scheint es demnach, als ob die Aktivitäten der *Twining*-Programme Staatsbeamten zu Gute kommen müssen, die bereits demokratische Verfahren in westlichen Demokratien persönlich erlebt haben« oder etwas anders formuliert »bedeutet dieser Befund vor allem, dass ein vorheriger direkter Kontakt mit demokratischen Prinzipien – unabhängig von ihrer Bewertung – notwendig zu sein scheint, damit ein späterer zielgerichteter Austausch über diese Prinzipien an bereits vorhandene Schemata, das heißt mentale Wissensstrukturen anknüpfen, diese aktivieren und deuten kann« (Freyburg 2011: 26). Wenn das Kooperationsprogramm tatsächlich erfolgreich an solche Wissensstrukturen anknüpfen könnte, dann wäre dies in der Tat ein Beleg für die Sozialisationsthese.

Diese Schlussfolgerung ist m.E. unbegründet, weil sie sich nicht schlüssig aus den Daten ableiten lässt. Der Befund eines signifikanten Interaktionsterms allein stützt Hypothese 2a nicht. Modell 3a enthält auch keine Variable, die es zuließe, eine Aussage über die Zurückweisung der Nullhypothese zu Hypothese 2b zu machen. Ein Vergleich der Niveaus der Einstellungen von Beamten mit und ohne Auslandserfahrungen deutet nicht auf ein »Anknüpfen« an frühere Erfahrungen hin; wenn überhaupt, dann eher auf eine Kompensation früherer negativer Erfahrungen. Vielmehr deuten die Daten auf das Vorliegen eines Selektionseffektes hin. Das sind freilich harsche Worte. Worin gründen meine Zweifel?

Zur Begründung meiner Befunde repliziere ich zunächst die berichteten Regressionsanalysen, wobei ich zur Schätzung der Parameter der linearen Modelle auf einen einfachen OLS-Schätzer zurückgreife (Tabelle 1). Die Autorin begründet die Verwendung des im Programm paket *MPlus* implementierten *maximum-likelihood*-Schätzverfahrens *MLMV* mit dessen »Robustheit gegenüber nicht-normalverteilten kontinuierlichen Daten« (Freyburg 2011: 13). Da für die Normalverteilung der Regressionskoeffizienten β lediglich die Normalverteilung der Störterme u vorgesetzt werden muss (und nicht die Normalverteilung der abhängigen Variablen oder Ähnliches) und für die Residuen der Regressionen die Normalitätsannahme nicht zu-

rückgewiesen werden kann, soll hier das einfachere OLS-Verfahren Anwendung finden.⁶

Tabelle 1: Replikation der Regressionsanalysen zum demokratischen Regieren (OLS-Schätzung)

	Model 1'	Model 2'	Model 3a'	Model 4'	Model 5'
Konstante	4.016*** (0.053)	4.012*** (0.092)	4.073*** (0.093)	4.032*** (0.089)	4.064*** (0.101)
Kooperation	0.084 (0.073)	0.068 (0.076)	-0.092 (0.098)		-0.054 (0.184)
Admin. Sozialisation		-0.036 (0.079)	-0.042 (0.077)	-0.045 (0.078)	-0.045 (0.078)
Auslandsaufenthalt		-0.052 (0.075)	-0.251* (0.108)	-0.049 (0.075)	-0.253* (0.109)
Ausl. Medien		-0.085 (0.098)	-0.055 (0.096)	-0.062 (0.095)	-0.037 (0.121)
Andere Programme		0.217** (0.074)	0.218** (0.072)	0.221** (0.074)	0.217** (0.073)
Koop. x Auslandsaufenthalt			0.364* (0.145)		0.366* (0.146)
Koop. x Ausl. Medien					-0.048 (0.196)
<i>N</i>	110	103	103	103	103
<i>R</i> ²	0.012	0.096	0.152	0.089	0.153
adj. <i>R</i> ²	0.003	0.050	0.099	0.052	0.090
Resid. sd	0.382	0.368	0.358	0.368	0.360

Standardfehler in Klammern; $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$

Wie Tabelle 1 zeigt, ergeben sich im Ergebnis nahezu identische Koeffizienten und Standardfehler. Es gibt drei Unterschiede zur ursprünglichen, Freyburgschen Tabelle 3. Erstens sind keine Schätzungen für Modell 3b abgebildet, das lediglich

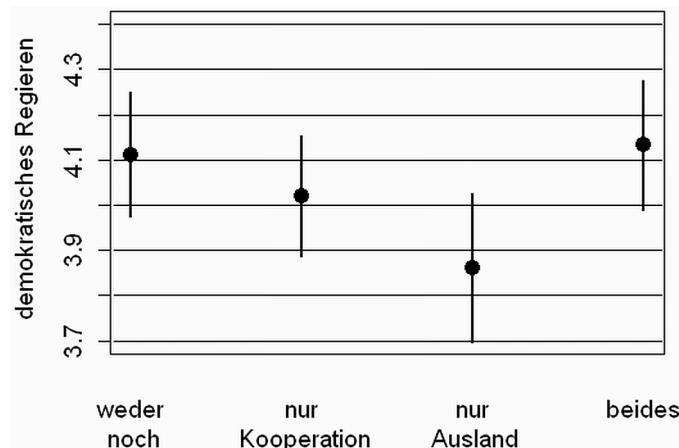
6 Etwas komplexer stellt sich die Sache bei der Regressionsanalyse der drei Einzeldimensionen dar. Grund hierfür ist der Generierung der Dimensionen aus zwei bzw. drei fünfstufigen Likertskalen, so dass je Dimension nur maximal acht Ausprägungen vorliegen (die dann als intervallskaliert interpretiert werden sollen). In der Folge kann die Normalverteilungsannahme hier nicht immer zurückgewiesen werden. Allerdings ist der OLS-Schätzer vergleichsweise robust gegen kleinere Verletzungen der Verteilungsannahmen, die zweitens durch monotone Transformationen der abhängigen Variablen behoben werden können.

eine lineare Transformation von Modell 3a darstellt,⁷ dafür aber ein weiteres Modell 5', das zum Test von Hypothese 2b einen weiteren Interaktionsterm zwischen den Variablen Kooperation und Ausländische Mediennutzung enthält. Erst Modell 5' erlaubt dann den Schluss, dass Hypothese 2b von den Daten nicht gestützt wird. Drittens berichtet Tabelle 1 Schätzungen für die Konstante der Regressionsgleichung, die in der ursprünglichen Tabelle wahrscheinlich wegen der Mittelwertstandardisierung der Variablen durch *MPlus* nicht erscheinen.

Eine mögliche Stützung von Hypothese 2a ergibt sich nun nicht aus dem Interaktionsterm von Kooperation und Auslandsaufenthalt in Modell 3a' (0,364), sondern dem signifikanten marginalen Effekt der Kooperation für Beamte mit Auslanderfahrung, oder $dY/dK = -0,092 + 0,364 = 0,272$ mit $p = 0,015$. Betrachtet man weiterhin die absoluten Effekte, dann ergibt sich folgendes Bild (Tabelle 1): Während die Kooperation bei Beamten mit Auslandserfahrung die Einstellung zu demokratischem Regieren von 3,87 auf 4,14 erhöht und bei Beamteninnen ohne Auslanderfahrung – allerdings nicht signifikant – reduziert, stimmen Beamte ohne Auslandserfahrung im Mittel dem Prinzip des demokratischen Regieren stärker zu als Beamteninnen mit Auslandserfahrung. Konkret gibt es keine Einstellungsunterschiede zwischen Beamten ohne Auslands- und Kooperationserfahrung und Beamteninnen, welche beide Erfahrungen haben. Wie kann das sein? Eine Interpretation ist, dass frühe Auslandsaufenthalte (zu Zwecken des Studiums) eher negative Einstellungen zu demokratischem Regieren hervorrufen, die dann in *Twining*-Projekten wieder revidiert werden können. Eine zweite Möglichkeit besteht im Vorliegen eines Selektionseffektes und zwar nicht bei der Auswahl der Projektteilnehmer, sondern schon der Rekrutierung der Beamten. Mit den vorliegenden Daten ist dies freilich nicht überprüfbar. In jedem Fall scheint das Hypothesen zwei zugrunde liegende Argument, dass nämlich Projektteilnehmer, die »bereits Erfahrungen mit demokratischem Regieren gemacht haben, eher dazu geneigt sind, ihre Einstellungen zu ändern, wenn sie erneut in Kontakt mit demokratischem Regieren kommen« (Freyburg 2011: 13) durch die Ergebnisse nicht gestützt.⁸

7 Entsprechend ergibt sich der Koeffizient für *Kooperation* in Freyburgs Modell 3b gerade als Summe der Koeffizienten von *Kooperation* und *Koop. x Auslandsaufenthalt* in Modell 3a (0,364 - 0,092 = 0,272).

8 Dass sich Beamte mit und ohne Auslandserfahrung in weiteren, eventuell relevanten Charakteristika unterscheiden, zeigt auch die Berücksichtigung der Variable »Geschlecht« in den Regressionen. Ersetzt man in Modell 3a' *Auslandserfahrung* mit *Männlich*, dann ergibt sich ein mit meiner Abbildung 1 nahezu identisches Bild der absoluten und marginalen Effekte.



Erhält man ein differenzierteres Bild, wenn man die einzelnen *Twinning*-Programme getrennt betrachtet? Die befragten Projektteilnehmer und -teilnehmerinnen hatten in einem von vier Programmen in den Bereichen Zölle, Wettbewerb, Gesundheit und Umwelt teilgenommen, welche sich hinsichtlich der Größe (beteiligte Beamte), der Dauer (9 bis 18 Monate) und dem Grad der Politisierung (niedrig bis hoch) unterscheiden (Freyburg 2011: 28). Für alle drei Faktoren lassen sich plausible Hypothesen hinsichtlich ihres Effektes auf die Einstellungsänderungen der Beamtinnen formulieren, die aber aufgrund der geringen Fallzahl von Programmen (nämlich vier) kaum zu testen sind. Tina Freyburg präsentiert dazu in Anhang IV eine größere Anzahl von einzelnen Regressionsanalysen für jedes *Twining*-Programm bei unterschiedlicher Spezifikation der abhängigen Variablen (drei konstituierende plus die übergeordnete Dimension demokratischen Regierens) und der unabhängigen Variablen.⁹ Dabei bilden für die Teilnehmer beispielsweise des *Twining*-Projekts im Bereich Wettbewerb ($n=10$) alle übrigen Befragten die Vergleichsgruppe, gleich ob sie an überhaupt keinem oder einem anderen *Twining*-Projekt teilgenommen haben. Wenn beispielsweise die Projektteilnahme stets einen positiven (und möglicherweise unterschiedlichen) Effekt auf die Einstellungen zum demokratischen Regieren hat, dann unterschätzt diese Vorgehensweise den Effekt des Wettbewerbsprojekts. Hinzu kommt, dass geschätzte Parameter zweier statistischer Modelle in der Regel nicht sinnvoll verglichen werden können. Eine einfache Lösung dieses Problems besteht in gleichzeitigen Schätzungen der Effekte der einzelnen *Twining*-Projekte in einem einzigen statistischen Modell. Tabelle 2 zeigt dies beispielhaft für die übergeordnete Dimension des demokratischen Regierens, wobei ich, mit dem Ziel die Hypothesen 2a und 2b zu testen, die Interaktionsterme der *Twining*-Projekte mit dem Auslandsaufenthalt und der ausländischen Medien-

9 Es werden offenbar nur solche Modelle präsentiert, die mindestens eine statistisch signifikante Parameterschätzung aufweisen.

nutzung berücksichtige. Die Spalten zeigen das Ergebnis der gewöhnlichen OLS-Regression mit approximativen Standardfehlern in runden Klammern sowie mittels *bootstrap* erzeugte Standardfehler in eckigen Klammern. Letztere werden aus der Varianz der geschätzten Koeffizienten ermittelt, nachdem die Modellschätzung mehrfach für Stichproben aus den vorliegenden Daten ausgeführt wurde (die Stichproben sind alle vom Umfang der Originaldaten und werden ohne Zurücklegen aus den Daten gezogen. Damit kommt man ohne spezifische Verteilungsannahmen aus, man nimmt lediglich an, die Grundgesamtheit verhalte sich zu den vorliegenden Daten wie die Daten zur *bootstrap*-Stichprobe).

Tabelle 2: Regressionsanalysen des demokratischen Regierens mit Twinning-Projekt-spezifischen Effekten

	Model 6	Model 7	Model 8
Konstante	4.001 (0.101)*** [0.294]	4.052 (0.093)*** [0.288]	4.061 (0.100)*** [0.291]
Koop.: Handel	0.065 (0.391) [0.427]	-0.315 (0.155)* [0.448]	-0.225 (0.456) [0.523]
Koop.: Wettbewerb	0.441 (0.386) [0.409]	-0.018 (0.221) [0.408]	0.096 (0.501) [0.434]
Koop.: Gesundheit	0.048 (0.237) [0.482]	-0.064 (0.137) [0.383]	0.003 (0.231) [0.480]
Koop.: Umwelt	-0.208 (0.389) [0.482]	-0.010 (0.136) [0.333]	-0.657 (0.432) [0.616]
Admin. Sozialisation	-0.023 (0.084) [0.286]	-0.031 (0.081) [0.284]	-0.044 (0.084) [0.292]
Auslandsaufenthalt	-0.053 (0.080)	-0.257 (0.107)*	-0.255 (0.108)*

	Model 6	Model 7	Model 8
	[0.298]	[0.352]	[0.350]
Ausl. Mediennutzung	-0.090 (0.124) [0.363]	-0.032 (0.101) [0.331]	-0.041 (0.122) [0.358]
Andere Programme	0.232 (0.076)** [0.286]	0.219 (0.073)** [0.284]	0.229 (0.074)** [0.288]
Koop.: Handel x Ausland		0.485 (0.250)† [0.527]	0.455 (0.271)† [0.550]
Koop.: Wettbewerb x Ausland		0.525 (0.314)† [0.426]	0.487 (0.348) [0.430]
Koop.: Gesundheit x Ausland		0.239 (0.189) [0.434]	0.271 (0.207) [0.469]
Koop.: Umwelt x Ausland		0.467 (0.223)* [0.521]	0.594 (0.237)* [0.523]
Koop.: Handel x Medien	-0.204 (0.418) [0.482]		-0.088 (0.442) [0.522]
Koop.: Wettbewerb x Medien	-0.248 (0.424) [0.403]		-0.113 (0.455) [0.353]
Koop.: Gesundheit x Medien	-0.003 (0.262) [0.512]		-0.097 (0.279) [0.542]

	Model 6	Model 7	Model 8
Koop.: Umwelt x Medien	0.400 (0.406) [0.435]		0.644 (0.418) [0.506]
<i>N</i>	103	103	103
<i>R</i> ²	0.162	0.216	0.241
adj. <i>R</i> ²	0.050	0.112	0.100
Resid. sd	0.368	0.356	0.358
OLS-Standardfehler in runden Klammern, *signifikant $p < .10$			
$*p < .05; **p < .01; ***p < .001;$			
Bootstrap-Standardfehler in eckigen Klammern (10000 Stichproben)			

Wie Tabelle 2 zeigt, sind die *bootstrap*-Standardfehler zumeist etwas größer als die approximativen Fehler. Modelle 6 und 7 berücksichtigen lediglich die Interaktionen der Kooperationsvariablen mit Medien bzw. Auslandsaufenthalt, Modell 8 beide Arten. Nach Modell 6 gibt es keine Evidenz für die in Hypothese eins und Hypothese 2b formulierten Erwartungen. Es gibt weder einen positiven Kooperationseffekt für Nutzer ausländischer Medien noch für Nichtnutzer. Einzig die Beteiligung an anderen Programmen hat den bekannten positiven Effekt. Für Modell 7 gilt Ähnliches. Hier zeigen die OLS-Standardfehler zwar signifikante Effekte für Teilnehmer in den *Twining*-Projekten Handel (negativer Effekt für Teilnehmer ohne Auslandserfahrung, was nicht zu Hypothese 2a passt) und Wettbewerb und Umwelt (jeweils positiver Effekt für Teilnehmer mit Auslandserfahrung, was im Einklang mit Hypothese 2a steht: $-0,018+0,525=0,507$ mit $p=0,02$ bzw. $-0,010+0,467=0,457$ mit $p=0,01$). Berücksichtigt man aber die durch *bootstrap* erhaltenen Standardfehler, dann erweist sich nur der Effekte für Wettbewerb als signifikant. Das entspricht nicht der Erwartung, »dass demokratische Sozialisation erleichtert wird, wenn die Interaktion in einem weniger politisierten Umfeld [z.B. Umwelt] stattfindet« (Freyburg 2011: 29). Der Befund, dass Teilnehmer »auch in politisierten Bereichen wie der Wettbewerbspolitik möglicherweise durch eine besonders konzentrierte Interaktion in demokratische Prinzipien und Praktiken sozialisiert werden« (Freyburg 2011: 29), wirft dann mehr Fragen auf, als er beantwortet. Modell 8 zeitigt ähnliche Ergebnisse. Die einzigen halbwegs robusten Effekte ergeben sich für die Kontrollvariable einer früheren Beteiligung an anderen Programmen, die die Einstellungen zum demokratischen Regieren positiv verändert, und einen früheren Auslandsaufenthalt, der scheinbar bei den Beamten negative Einstellungen zur Demokratie hervorgerufen und hinterlassen hat.

5. Zusammenfassung

Kann Demokratie gelernt werden? Können Menschen, die ein halbes Leben in autoritären Strukturen sozialisiert wurden, von demokratischen Prinzipien eines partizipatorischen und transparenten Regierens überzeugt werden, auch wenn und gerade wenn dies ihre unmittelbare Handlungsfreiheit und Macht einschränkt? Wir wünschen es, aber unser Alltagswissen lehrt uns auch, dass Menschen oft sich Regeln unterordnen und diese befolgen, wenn es nur sozial erwünscht erscheint. Auch sind Einstellungsänderungen langfristige Prozesse, die einer Vielzahl von Stimuli und Situationen bedürfen, die alte Einstellungen in Fragen stellen und die Möglichkeit des Abwägens der alten mit der neuen Ordnung bieten. Insofern ist es nicht verwunderlich, dass Arbeiten zur demokratischen Sozialisation von Individuen selten mit klaren Befunden aufwarten können. Tina Freyburg hat vor diesem Hintergrund ein kluges Design zur Untersuchung der demokratischen Sozialisation von Staatsbeamten in autoritären Systemen gewählt. Mit einer anonymen *Survey*-Befragung der Beamten wurden Einstellungen erhoben anstelle Verhalten beobachtet. Der Vergleich von Beamten mit und ohne Beteiligung an einem *Twining*-Projekt der EU sollte in einem quasi-experimentellen Design den positiven Effekt der demokratischen Sozialisation aufzeigen. Die Evidenz hierfür ist allerdings vage. Was können wir daraus für zukünftige Studien lernen? Ich sehe hier drei Dinge, die sich auf die Notwendigkeit einer konsequenten Umsetzung eines experimentellen Designs, der Wichtigkeit der Erhebung von individuellen Merkmalen in derartigen *Surveys*, und auf die Methoden der Messung von Einstellungen darin beziehen.

Dazu hilft es meines Erachtens, sich über die Bedingungen von kausalen Schlüssen klar zu werden. Kausale Inferenz bedarf stets einer kontrafaktischen Annahme. Wenn wir sagen, dass ein Faktor X einen Einfluss auf Y hat, dann meinen wir damit, dass sich der potentielle Wert von Y unterscheidet, je nachdem, ob X , in der einfachsten Form, vorliegt (1) oder nicht (0) (Holland 1986). Da für ein einzelnes Individuum i , X_i empirisch nur einen der beiden Zustände annehmen kann, können wir auch nur einen der beiden potentiellen Werte $Y_i(0)$ bzw. $Y_i(1)$ beobachten. Der kausale Effekt auf der Ebene der einzelnen Beobachtungseinheit, also $Y_i(1) - Y_i(0)$, ist prinzipiell nicht beobachtbar. Das ist das grundlegende Problem der kausalen Inferenz.¹⁰ In der Forschungspraxis betrachtet man deshalb den durchschnittlichen Effekt eines Faktors oder *treatments*, den man aus dem Vergleich der Werte einer größeren Anzahl von Individuen, bei denen der Faktor vorliegt oder nicht, gewinnt. Voraussetzung für eine zulässige kausale Inferenz ist dann allerdings, dass sich die beiden Gruppen sonst im Durchschnitt nicht unterscheiden. Das trifft auf experimentelle Designs zu, bei denen das *treatment* zufällig zugewiesen wird, wie bei Labor-, Feld- und Survey-Experimenten oder die Zuweisung nach einem Mechanismus.

10 Ein zweites besteht in der theoretischen Begründung des kausalen Mechanismus (vgl. Imai et al. 2011).

mus erfolgt, der – nach allem was uns plausibel erscheint – unabhängig von Y ist, wie in natürlichen Experimenten.

Für viele Fragestellungen sind wir auf Beobachtungsstudien angewiesen, in denen wir für Unterschiede in den (*pre-treatment*) Kovariaten zwischen *treatment*- und Kontrollgruppe kontrollieren müssen, etwa durch Regressions- oder *matching*-Verfahren. Eine notwendige Voraussetzung für kausale Inferenz ist, dass wir plausibel argumentieren können, dass es keine weiteren Variablen gibt, die sowohl das *treatment* als auch die abhängige Variable beeinflussen. In diesen Fällen gilt es bei der Erhebung der Daten sicherzustellen, dass hinreichend individuelle Merkmale vorliegen, die für eine Selektion in die *treatment*-Gruppe wichtig gewesen sein könnten und potentiell einen Einfluss auf die abhängige Variable haben – im vorliegenden Fall etwa Bildung und beruflicher Status. Eine weitere Schlussfolgerung betrifft die Messung von Einstellungen zu sensiblen Fragen, bei denen sozial erwünschtes Antwortverhalten angenommen werden muss (vgl. Holbrook 2011). Insofern, wie in der vorliegenden Studie, die Unterschiede zwischen Gruppen von Individuen von Interesse sind, spielt eine durchschnittliche Erhöhung des Niveaus der Einstellung keine Rolle. Dies allerdings setzt voraus, dass sich das Ausmaß an sozial erwünschtem Antwortverhalten nicht zwischen den beiden Gruppen unterscheidet, was oftmals nicht überprüft werden kann. Als Ausweg bietet sich hier beispielsweise die Listentechnik an, bei der Individuen nur die Anzahl von Zustimmungen zu einer ganzen Liste von teilweise sensiblen Fragen angeben müssen, nicht aber, ob sie einem bestimmten Item zustimmen. Das Ausmaß an sozial erwünschtem Antwortverhalten lässt sich dann über den Vergleich von zwei Gruppen in einem *Survey*-Experiment bestimmen, welchen die sensible(n) Frage(n) mit bzw. nicht mit auf die Liste gesetzt wurde (vgl. Kuklinski/Cobb 1998). Diese und andere Techniken können uns helfen, Einstellungsänderungen, gerade wenn sie sensible Themenbereiche berühren, nicht nur experimentell im Labor zu studieren, sondern in der natürlichen Umwelt der Betroffenen, der Bürger, Beamtinnen und Politiker, in Europa wie in Marokko. Kurzum, auch wenn die Fallstricke, die bei einer sorgfältigen empirisch-quantitativen Analyse lauern, zahlreich sind: Wenn wir die kausalen Mechanismen einer demokratischen Sozialisation nachweisen wollen, führt wohl kein Weg an einem intelligenten Einsatz von experimentellen Verfahren in natürlichen Umgebungen vorbei. Tina Freyburgs Studie ist ein wichtiger Beitrag hierzu.

Literatur

- Eagly, Alice/Chaiken, Shelly 1993: The Psychology of Attitudes. Orlando, FL.
- Freyburg, Tina 2009: Democrats without Democracy? Linkage and Socialization into Democratic Governance in Authoritarian Regimes (NCCR Democracy Working Paper 37), Zürich.
- Freyburg, Tina 2011: Demokratisierung durch Zusammenarbeit? Funktionale Kooperation mit autoritären Regimen und Sozialisation in demokratischem Regieren, in: Zeitschrift für Internationale Beziehungen 18: 1, 5-46.

- Gelman, Andrew/Hill, Jennifer* 2009: Data Analysis Using Regression and Multilevel/Hierarchical Models, Cambridge.
- Holbrook, Allyson L.* 2011: Attitude Change Experiments in Political Science, in: Druckman, James N./Green, Donald P./Kuklinski, James H./Lupia, Arthur (Hrsg.): Cambridge Handbook of Experimental Political Science, Cambridge, MA, 141-154.
- Hooghe, Liesbet* 2005: Several Roads Lead to International Norms, but Few via International Socialization: A Case Study of the European Commission, in: International Organization 59: 4, 861-898.
- Kuklinski, James H./Cobb, Michael D.* 1998: When White Southerners Converse About Race, in: Hurwitz, Jon/Peffley, Mark (Hrsg.): Perception and Prejudice, New Haven, MA, 35-57.
- Searing, Donald/Wright, Gerald/Rabinowitz, George* 1976: The Primacy Principle: Attitude Change and Political Socialization, in: British Journal of Political Science 6: 1, 83-113.