



Projektbericht

RWI – Leibniz-Institut für Wirtschaftsforschung

**Meta-Analyse kausaler Effekte von
verhaltensökonomischen Interventionen auf
den Energieverbrauch privater Haushalte**

**Projektbericht an das Ministerium für Ländlichen Raum
und Verbraucherschutz Baden-Württemberg und das
Forschungszentrum Verbraucher, Markt und Politik |
CCMP an der Zeppelin Universität**

28. Dezember 2017



Impressum

Herausgeber:

RWI – Leibniz-Institut für Wirtschaftsforschung
Hohenzollernstraße 1-3 | 45128 Essen, Germany

Postanschrift:

Postfach 10 30 54 | 45030 Essen, Germany

Fon: +49 201-81 49-0 | E-Mail: rwi@rwi-essen.de
www.rwi-essen.de

Vorstand

Prof. Dr. Christoph M. Schmidt (Präsident)

Prof. Dr. Thomas K. Bauer (Vizepräsident)

Dr. Stefan Rumpf

© RWI 2017

Der Nachdruck, auch auszugsweise, ist nur mit Genehmigung des RWI gestattet.

RWI Projektbericht

Schriftleitung: Prof. Dr. Christoph M. Schmidt

Gestaltung: Daniela Schwindt, Magdalena Franke, Claudia Lohkamp

Meta-Analyse kausaler Effekte von verhaltensökonomischen Interventionen auf den Energieverbrauch privater Haushalte

Projektbericht an das Ministerium für Ländlichen Raum und Verbraucherschutz Baden-Württemberg und das Forschungszentrum Verbraucher, Markt und Politik | CCMP an der Zeppelin Universität

28. Dezember 2017

Projektteam

Dr. Mark Andor (Leitung), Dr. Gunther Bensch (Leitung), Katja Fels und Nadine Kneppel

Das Projektteam dankt den Mitarbeitern des Forschungszentrums Verbraucher, Markt und Politik (CCMP) an der Zeppelin Universität für die gute Zusammenarbeit und viele hilfreich Kommentare zu diesem Bericht, insbesondere Sabine Bietz, Micha Kaiser, Julius Rauber und Lucia Reisch. Wir danken Marvin Gleue, Mandy Kurth, Charlotte Plinke, Heiko Röhrs, Lara Roetzel, Fabian Streib, Marc Teipel und Maximilian Zettler für die hervorragende Unterstützung.

RWI – Leibniz-Institut für Wirtschaftsforschung

Meta-Analyse kausaler Effekte von verhaltensökonomischen Interventionen auf den Energieverbrauch privater Haushalte

**Projektbericht an das Ministerium für Ländlichen
Raum und Verbraucherschutz Baden-Württemberg
und das Forschungszentrum Verbraucher, Markt und
Politik | CCMP an der Zeppelin Universität**

28. Dezember 2017

Inhaltsverzeichnis

1.	Einleitung	3
2.	Verhaltensökonomische Interventionen zur Förderung von Energieeinsparungen	4
2.1.	Verhaltensökonomische Interventionen als Teil des energiepolitischen Instrumentariums	4
2.2.	Typen verhaltensökonomischer Interventionen zur Förderung von Energieeinsparungen	6
2.3.	Vorausgehende wissenschaftliche Übersichtsarbeiten zur Thematik.....	10
3.	Methoden	12
3.1.	Kriterien für die Einbeziehung von Studien in die Meta-Analyse.....	12
3.2.	Systematische Suche relevanter Primärliteratur.....	13
3.3.	Datensammlung.....	15
3.4.	Effektgrößenberechnung.....	15
4.	Resultate	18
4.1.	Ergebnisse der systematischen Suche	18
4.2.	Finale Auswahl von Effektschätzern und Studien.....	19
4.3.	Studienbeschreibung	20
4.4.	Synthese der quantitativen Ergebnisse	24
4.4.1.	Feedback.....	24
4.4.2.	Sozialer Vergleich.....	28
4.4.3.	Zielsetzung und Selbstbindung	31
4.4.4.	Labeling.....	34
4.5.	Moderatoranalyse	37
5.	Zusammenfassung, Schlussfolgerungen und Handlungsempfehlungen.....	41
5.1.	Zusammenfassung	41
5.2.	Schlussfolgerungen.....	42
5.3.	Handlungsempfehlungen	43
6.	Literatur	46
6.1.	Zitierte Literatur.....	46
6.2.	In die Meta-Analyse einbezogene Literatur	52
A	Anhang.....	57
A.1	Schlagwortkombinationen für die Datenbanksuche	57
A.2	Liste der extrahierten Daten.....	59
A.3	Hierarchisierung von Schätzern und Standardabweichungen	62
A.4	Berechnungsformeln der Effektgrößenstatistiken	63
A.5	Anleitung zum Lesen und Interpretieren der <i>Forest Plots</i>	65

Verzeichnis der Tabellen und Schaubilder

Tabelle 1	Klassifizierung von Interventionstypen	5
Tabelle 2	Internalitäten und ausgewählte verhaltensökonomische Interventionen	7
Tabelle 3	Bewertungsfragen als Grundlage für die Einbeziehung der Primärliteratur.....	14
Tabelle 4	Übersicht über Kategorien und beispielhafte Artikel.....	14
Tabelle 5	Zur Zusammenfassung von Effektgrößen herangezogene Dimensionen.....	15
Tabelle 6	Artikel-Studien-Kombinationen und beispielhafte Artikel	19
Tabelle 7	Weitere Ausschlüsse von Studien und Effektschätzern	20

Tabelle 8 Zusammenfassung der betrachteten Studien	21
Tabelle 9 Eigenschaften der Datenerhebung der Studien	22
Tabelle 10 Weitere Studieneigenschaften	22
Tabelle 11 Metaregressionsergebnisse, alle Interventionen und Ergebnisvariablen gepoolt.....	38
Tabelle A.1 Schlagwortkombinationen für die Datenbanksuche	57
Tabelle A.2 Variablenliste.....	59
Tabelle A.3 Hierarchisierung von Schätzern, die zu gepoolten Effektschätzern beitragen	62
Tabelle A.4 Verwendete standardisierte Effektgrößenstatistiken.....	63
Tabelle A.5 Hierarchisierung von Standardabweichungen der Ergebnisvariable zur Berechnung der standardisierten Effektgröße	64
Abbildung 1 Überblick über den Such- und Selektionsprozess.....	18
Abbildung 2 Anzahl der Studienteilnehmer	21
Abbildung 3 Interventionskombinationen und Ergebnisvariablen nach Intervention.....	24
Abbildung 4 Effekt von Feedback auf den Stromverbrauch	26
Abbildung 5 Effekt von Feedback auf den Gasverbrauch	27
Abbildung 6 Effekt von Feedback auf den Wasserverbrauch	27
Abbildung 7 Effekt von Sozialen Vergleichen auf den Stromverbrauch	28
Abbildung 8 Effekt von Sozialen Vergleichen auf den Gasverbrauch	29
Abbildung 9 Effekt von Sozialen Vergleichen auf den Wasserverbrauch	29
Abbildung 10 Effekt von selbstbestimmter Zielsetzung und Selbstbindung auf den Stromverbrauch	32
Abbildung 11 Effekt von fremdbestimmter Zielsetzung auf den Stromverbrauch	33
Abbildung 12 Effekt von Zielsetzung und Selbstbindung auf den Gasverbrauch	33
Abbildung 13 Effekt von Zielsetzung und Selbstbindung auf den Wasserverbrauch	34
Abbildung 14 Quantifizierung der Effekte von Labeling unter realen Studienbedingungen.....	35
Abbildung 15 Quantifizierung der Effekte von Labeling in hypothetischen Entscheidungssituationen.....	36
Abbildung 16 Vergleich der Effekte in der kurzen und längeren Frist	40

1. Einleitung

Weltweit stehen Klimaschutzprogramme auf der politischen Agenda. Als ein Mittel um die ambitionierten Klimagasreduktionen zu erreichen, verfolgen die Europäische Union, Deutschland und viele weitere Nationen weltweit das Ziel, den Energieverbrauch drastisch zu reduzieren. Im Rahmen der „Energiewende Baden-Württemberg“ hat sich die Landesregierung das ehrgeizige Ziel gesetzt, bis 2050 einen 50 Prozent geringeren Verbrauch als im Jahr 2010 zu erreichen. Es stellt sich jedoch die grundsätzliche Frage, wie solch starke Energieverbrauchsreduktionen erzielt werden können.

Vor diesem Hintergrund ist die Politik in zahlreichen Industrieländern zunehmend an Maßnahmen interessiert, welche private Haushalte zum Energiesparen bewegen. An Aufmerksamkeit gewinnen dabei insbesondere verhaltensökonomische Interventionen. Mit der Vergabe des Nobelpreises an den Ökonomen Richard Thaler wurde jüngst ein Vordenker dieses unter dem Begriff „Nudging“ (engl. für Stups oder Schubs) bekannt gewordenen Ansatzes ausgezeichnet. In dem gleichnamigen Buch definiert Thaler Nudges als Aspekte der Entscheidungsstruktur, welche das Verhalten der Menschen in vorhersehbarer Weise verändern, ohne eine Handlungsoption zu verbieten oder deren ökonomische Anreizstrukturen signifikant zu verändern (Thaler und Sunstein 2009). Die Attraktivität von Nudges liegt vor allem darin, dass sie in der Regel kostengünstig umgesetzt werden können und gleichzeitig weniger in die Wahlmöglichkeiten der Verbraucher eingreifen als beispielsweise Steuern oder Verbote. Diese Eigenschaften machen verhaltensökonomische Maßnahmen für die Politik zu einem attraktiven und potentiell kosteneffizienten Ansatz bei der Senkung des Energieverbrauchs (Allcott 2015). Sollten sich die Maßnahmen als dauerhaft und in unterschiedlichen Kontexten wirksam erweisen, könnten sie als integraler und komplementärer Bestandteil der Klimapolitik etabliert werden (Allcott und Mullainathan 2010). Vermehrt erforschen Wissenschaftler daher die Effekte und Auswirkungen nicht-monetärer Interventionen auf den privaten Energieverbrauch.

Die vorliegende Studie fasst die vorhandene Forschung zu Effekten von verhaltensökonomischen Interventionen auf das Energiesparverhalten privater Haushalte systematisch zusammen und präsentiert die Ergebnisse in Form einer Meta-Analyse. Eine solche Meta-Analyse wendet quantitative und statistische Methoden an, um allgemeine Effektgrößen von Interventionen anhand der Einzelergebnisse unterschiedlicher Primärstudien zu berechnen. Konkret werden fünf der am weitesten verbreiteten und als politisch machbar einzuschätzenden verhaltensökonomischen Interventionen analysiert: Feedback (Rückmeldung), Sozialer Vergleich, Selbstbindung, Zielsetzung sowie Labeling.

Unsere Studie erweitert eigene Arbeiten zu dem Thema (Andor und Fels 2017a) und baut auf früheren Reviews anderer Autoren auf. Diese älteren Überblicksstudien beschränken sich auf Teile der genannten Interventionen. Labeling wurde beispielsweise bislang nicht systematisch untersucht, obwohl Energie-Labels weltweit bereits im großen Umfang eingesetzt werden und potenziell Millionen von Kaufentscheidungen beeinflussen. Die Studien weisen zudem auf Probleme bei der Einbeziehung von Effekten aus Studien hin, die aufgrund methodischer Unzulänglichkeiten lediglich in der Lage sind eine Gleichzeitigkeit von Entwicklungen abzubilden (Korrelation) anstatt kausale Effekte zu identifizieren, also die ursächliche Wirkungsbeziehung zwischen einer Maßnahme und der daraus resultierenden Verhaltensänderung. So finden die Autoren Anzeichen dafür, dass die Effekte mit der Qualität des angewandten Evaluierungsansatzes abnehmen: Ansätze, die besser in der Lage sind kausale von nicht-kausalen Effekten zu trennen, doku-

mentieren tendenziell eine geringere Wirksamkeit der Interventionen. Für die politische Relevanz von Studienergebnissen ist die methodische Qualität somit entscheidend (Imbens und Wooldridge 2009).

Die vorliegende Studie greift dieses Defizit in der bisherigen Forschung auf, indem ausschließlich Studien betrachtet werden, die eine empirische Methode verwenden, welche eine Identifizierung kausaler Effekte zwischen der Intervention und dem Energieverbrauch ermöglicht. Unseres Wissens nach ist dies die erste Meta-Analyse, die systematisch alle publizierten Ergebnisse der Verhaltensökonomik sowie verwandter Forschungsgebiete auf Basis einer rigorosen Bewertung der zugrundeliegenden Methodik zusammenstellt und analysiert. Der Fokus dieser Meta-Analyse liegt auf der Berücksichtigung von Studien, die Effekte der Maßnahmen auf den Energieverbrauch, konkret den Strom- und Gasverbrauch, untersuchen. Zusätzlich werden Studien berücksichtigt, welche die Effekte auf den Wasserverbrauch ermitteln, da die Maßnahmen auch in diesem Fall zum Zwecke der Reduzierung des Ressourcenverbrauchs angewendet werden und die Grundmechanismen als ähnlich einzuschätzen sind. Um einen möglichst umfassenden und aktuellen Überblick über die relevante ökonomische und psychologische Forschung zu geben, berücksichtigt die Meta-Analyse alle Studien, die bis Mai 2017 als Diskussionspapiere oder in begutachteten Fachzeitschriften veröffentlicht wurden. Dies ist umso wichtiger, als es in der jüngeren Vergangenheit eine Zunahme an qualitativ hochwertigen Studien gab, insbesondere in Form großangelegter randomisierter Feldexperimente.

Im Folgenden werden die in diesem Ergebnisbericht vorgestellten verhaltensökonomischen Interventionen eingeführt, auch in Abgrenzung zu anderen Instrumenten zur Förderung von Energieeinsparungen (Kapitel 2), bevor die angewendete Methodik skizziert wird (Kapitel 3). Kapitel 4 stellt die Analysen und deren Resultate vor, bevor Kapitel 5 mit einer zusammenfassenden Diskussion abschließt, die auch auf die Frage eingeht, welche Handlungsempfehlungen sich für die baden-württembergische Verbraucherpolitik ableiten lassen.

2. Verhaltensökonomische Interventionen zur Förderung von Energieeinsparungen

2.1. Verhaltensökonomische Interventionen als Teil des energiepolitischen Instrumentariums

In Industriestaaten ist der Energieverbrauch des privaten Sektors für einen bedeutenden Anteil der jährlichen Emissionen verantwortlich. Um diesen zu drosseln, stehen der Politik verschiedene Instrumente zur Verfügung, die in unterschiedlichem Ausmaß die Entscheidungen des Einzelnen zu beeinflussen versuchen. Tabelle 1 stellt diese entsprechend ihrer Auswirkung auf Wahlmöglichkeiten dar. Nudging wird dadurch im Maßnahmenportfolio abgegrenzt von alternativen Ansätzen, die von grundsätzlichen Verboten wie einem allgemeinen Sonntagsfahrverbot bis zur bloßen Bereitstellung von Informationen, z.B. in Form von Energiespartipps, reichen.

Nudges erweitern somit den politischen Instrumentenkasten um eine dritte Wahloption: Neben „harten“ Eingriffen mittels Verboten und Geboten sowie „mittleren“ Eingriffen mittels finanziellen Anreizen (Steuern und Subventionen) ermöglichen Nudges einen so genannten „weichen“ bzw. sanften oder libertären Paternalismus (Thaler und Sunstein 2003; Kirchgässner 2017). Dieser zeichnet sich dadurch aus, dass er etwa über die Änderung der Voreinstellung Anpassungen der Entscheidungsarchitektur vornimmt, dabei die Wahlmöglichkeiten der Individuen jedoch nicht einschränkt. Das bekannteste Beispiel für eine geänderte Voreinstellung ist die Organspenderegulation. In manchen Ländern werden nur diejenigen Bürger Organspender, die sich aktiv melden (Opt-in), während in anderen Ländern jeder prinzipiell als Organspender geführt wird, so

lange er sich nicht aus dem Register abmeldet (Opt-out) (Thaler und Sunstein 2008). Im Energiebereich könnten Änderungen der Voreinstellung beispielsweise in Bezug auf die Wahl eines Stromtarifs mit erneuerbaren Energien eingesetzt werden (Sunstein und Reisch 2014).

Tabelle 1
Klassifizierung von Interventionstypen

Gesetze und Regulierungen		Fiskalische Maßnahmen		Nicht-regulatorische, nicht-fiskalische Maßnahmen			
grundsätzliche Verbote	Beschränkungen und Regulierungen	Fiskalische negative Anreize	Fiskalische positive Anreize	Nicht-fiskalische Anreize	Nudging	Überzeugungsarbeit	Bereitstellung von Informationen
<i>Allgemeines Sonntagsfahrverbot</i>	<i>Energetische Mindeststandards</i>	<i>Mineralölsteuer</i>	<i>Befreiung von Elektroautos von der Kraftfahrzeugsteuer</i>	<i>Kampagne „Mit dem Rad zur Arbeit“</i>	<i>Siehe Tabelle 2</i>	<i>Stromwechsel-Kampagnen</i>	<i>Energiespartipps</i>
Wahlmöglichkeit ausschließen	Wahlmöglichkeit beschränken	Wahlmöglichkeit ermöglichen und leiten					

Quelle: Eigene Darstellung basierend auf House of Lords (2011).

Ob mehr Nudges als Politikinstrumente eingesetzt werden sollten, ist jedoch umstritten. Der häufigste Einwand lautet, dass dieser „weiche“ staatliche Eingriff im Gegensatz zu Steuern oder Verboten von den Bürgern häufig nicht als solcher erkannt wird (Kasperbauer 2017). Durch die zunehmende Popularität bestehe zudem die Gefahr, dass immer mehr Nudges zur Anwendung kommen, die in Konkurrenz zueinander treten und gegenseitig ihre Effekte abschwächen oder aufheben (Bovens 2009; Baldwin 2014). Hansen und Jespersen (2013) merken zudem an, dass es bislang wenig Erfahrung mit diesen auf unbewusste kognitive Prozesse ausgerichteten Politikmaßnahmen durch Regierungen gibt und dass solche Eingriffe zu neuen ethischen Problemen führen können. So stellen manche Kritiker die Frage, warum Politiker, die selbst auch durch Biases zu systematischen Fehlentscheidungen verleitet werden können, besser wissen sollten, was das Beste für die Bürger ist als die Bürger selbst. Ein „idealer“ Entscheidungsarchitekt müsse nicht nur die wahren Bedürfnisse der Menschen kennen, sondern auch wissenschaftliche Erkenntnisse verstehen und einordnen können sowie keine konfligierenden Interessen haben (Gigerenzer 2015). Darüber hinaus kosten Nudges zwar oft wenig auf der Seite der Ausführenden, ob jedoch auf Seite der „Genudgden“ (psychologische) Kosten etwa durch Stress, ein schlechtes Gewissen oder als unangenehm empfundenen Entscheidungsdruck entstehen und wie hoch diese sind, darüber gibt es erst wenige Erkenntnisse (siehe als erste wegweisende Studien zu diesem Thema Allcott und Kessler 2017 sowie Damgaard und Gravert 2018).

Befürworter des „weichen Paternalismus“ argumentieren hingegen, dass Entscheidungen immer unter Umständen außerhalb der eigenen direkten Kontrolle getroffen werden und Nudges lediglich die Gesamtheit der Faktoren verändern, die in diese Entscheidungen einfließen. Das Gefühl, die aktuell vorherrschende Entscheidungsarchitektur beinhalte keine (staatliche) Beeinflussung, entstehe nur deswegen, weil eben jene Voreinstellung als selbstverständlich gegeben hingenommen werde. Allein durch ihre Existenz beeinflusse sie aber die Präferenzen der Menschen (Thaler und Sunstein 2003). Die Annahme der Nudging-Kritiker, dass Menschen hinreichend rational und unbeeinflusst von äußeren Kräften agieren, sei darüber hinaus unrealistisch (Kasperbauer 2017). Ein gewisses Maß an Nudges ist laut Sunstein (2015) im öffentlichen Sektor unausweichlich. Insbesondere Energiegewinnung und Energiekonsum sind nach Einschätzung von Kasperbauer (2017) bereits so „massively architected“, d.h. durch äußere Faktoren (z.B. soziale Normen und Energieinfrastruktur) bestimmt, dass Nudges keine zusätzliche Beschränkung der Autonomie des Einzelnen bedeuten würden.

Dass es einen „idealen“ Entscheidungsarchitekten geben muss, halten Befürworter des Nudge-Ansatzes in der Politik zudem nicht für notwendig. Sie argumentieren, dass der demokratische Prozess ungewollte Manipulation verhindern werde, so lange alle eingesetzten politischen Maßnahmen transparent gemacht würden (Kirchgässner 2014; Purnhagen und Reisch 2016; Sunstein 2017b). Darüber hinaus zeigen Studien über mehrere Länder hinweg eine positive Einstellung der Bürger gegenüber der Anwendung von Nudges, so lange etwa geänderte Voreinstellungen nicht als konträr zur Präferenz der Mehrheit der Bevölkerung empfunden werden (Sunstein 2017a; Reisch und Sunstein 2016; Sunstein et al. 2017). Viele Wissenschaftler sehen einen großen Vorteil der Integrierung von verhaltenswissenschaftlichen Erkenntnissen in den Politikprozess darin, dass dieser effizienter und zielgenauer werde (Ciriolo 2011; Madrian 2014; Chetty 2015; Purnhagen und Reisch 2016).

Bei der Abwägung, ob in einem speziellen Fall Nudges als Politikinstrument eingesetzt werden, sollten daher vor allem drei Aspekte Berücksichtigung finden: (i) Unter welchen Umständen liegt ein Verhalten vor, das aus Wohlfahrtssicht eines Nudges bedarf? (ii) Was sind die wahren Wünsche und Bedürfnisse der Menschen, die genudged werden sollen? Wie homogen oder inhomogen sind diese zwischen den Einzelnen? Ist die gewählte Nudge geeignet, um diese Bedürfnisse zu adressieren? (iii) Wollen die Betroffenen genudged werden? Was sind die Kosten, die auf Seiten der Nudgenden und Genudgeden entstehen und in welchem Verhältnis stehen diese zu dem erwarteten Effekt (vgl. auch Bosworth und Bartke 2014) Die Antworten auf diese Fragen sind nicht trivial und bedürfen z.T. weiterer empirischer Untersuchungen, bieten aber einen guten Leitfaden für die Anwendung von Nudges im politischen Kontext.

2.2. Typen verhaltensökonomischer Interventionen zur Förderung von Energieeinsparungen

Verbraucher können auf zwei Wegen Energie einsparen: Zum einen durch Änderungen ihres Konsumverhaltens, indem sie beispielsweise weniger heizen. Zum anderen durch Anpassung ihres Kaufverhaltens hin zu mehr Investitionen in Energieeffizienz, etwa wenn sie eine energieeffiziente Waschmaschine erwerben.¹ Mit politischen Rahmenbedingungen können sowohl die Kaufentscheidung als auch das Konsumverhalten gezielt beeinflusst werden. Während stärkere politische Eingriffe wie Verbote oder Steuern in der Regel mit negativen Externalitäten wie Verschmutzung oder Klimaschäden gerechtfertigt werden, dienen staatliche Eingriffe durch verhaltensökonomische Maßnahmen im Energiebereich zwar ebenfalls dazu, Externalitäten entgegenzuwirken. Sie werden darüber hinaus jedoch vor allem mit so genannten Internalitäten gerechtfertigt, womit kognitive Verzerrungen gemeint sind, die – gemessen am individuell erfahrenen Nutzen des Individuums – optimale Entscheidungen verhindern (Chetty 2015).

In Bezug auf Energieeffizienz sind laut Allcott (2016) vor allem sechs Internalitäten für Entscheidungen von Konsumenten verantwortlich, die als nicht-optimal angesehen werden können: Present Bias (zu starke Gegenwartstendenz), Bias toward Concentration (Konzentrations-Bias), Biased Beliefs (systematisch verzerrte Einschätzungen), kostenintensive Informationsbeschaffung sowie exogene und endogene Unachtsamkeit (vgl. auch Reisch und Sunstein 2017). Im Fokus der vorliegenden Meta-Analyse stehen jene verhaltensökonomischen Interventionen, die am weitesten verbreitet und am besten dazu geeignet sind, diesen Internalitäten entgegen zu wirken (siehe Tabelle 2). Die sich daraus ergebenden fünf ausgewählten Nudges und die Internalitäten, die sie adressieren, werden im Folgenden näher erläutert.

¹ Der Kauf eines energieeffizienten Geräts wird zu geringerem Energieverbrauch führen, wenn die Effizienzsteigerung nicht vollständig von einer – durch die Effizienzsteigerung hervorgerufenen – erhöhten Nutzung des Geräts (auch Reboundeffekt genannt, siehe Frondel und Vance 2013) aufgewogen wird.

Tabelle 2

Internalitäten und ausgewählte verhaltensökonomische Interventionen

Internalität	Gewählte Interventionstypen
Gegenwartstendenz	Selbstbindung, Zielsetzung
Konzentrations-Bias	Labeling
Biased Beliefs (verzerrte Einschätzung des eigenen Verbrauchsverhaltens)	Feedback, Sozialer Vergleich, Labeling
kostenintensive Informationsbeschaffung	Feedback, Labeling
exogene und endogene Unachtsamkeit	Labeling

Quelle: Eigene Darstellung.

Feedback

Im Alltag erfahren Verbraucher für gewöhnlich nur selten etwas über ihren Energieverbrauch. Während sie jederzeit die Möglichkeit haben, ihren Verbrauchsstand abzufragen, ist eine solche Anfrage mit *kostenintensiver Informationsbeschaffung* (insbesondere in Form von Opportunitätskosten) verbunden. Während es beispielsweise häufig theoretisch möglich ist, täglich oder gar stündlich den Stromverbrauch des eigenen Haushalts abzulesen, indem man im Keller den Stromzähler abliest, findet dies im Allgemeinen eher nicht statt. Da der Energieverbrauch in Einheiten wie Kilowattstunden oder Megajoule abstrakt und nicht direkt erfahrbar ist (Karlin et al. 2015), ist es für den Einzelnen zudem schwierig, bewusst Energie zu verbrauchen, selbst wenn er sich beispielsweise als umweltfreundlichen Verbraucher versteht. Verbraucher können somit *verzerrten Einschätzungen* des eigenen Verbrauchsverhaltens (*Biased Beliefs*) unterliegen, sowohl im Allgemeinen als auch hinsichtlich bestimmter Geräte.

Feedbackinterventionen versuchen, diesen Internalitäten entgegenzuwirken. Allgemein werden sämtliche Maßnahmen als Feedback bezeichnet, die einem Akteur niedrigschwellig Rückmeldung zu seinem Verbrauchsverhalten geben. Feedback verknüpft damit Handeln direkt mit einer Konsequenz (Madrian 2014). Die Empfänger von Feedback können diese Informationen nutzen, um ihr zukünftiges Verhalten in Richtung eines gewünschten Ergebnisses anzupassen (Karlin et al. 2015). Im Kontext von Energiekonsum bezieht sich Feedback vor allem auf zeitnahe Informationen über konkrete Energieverbrauchswerte von Einzelpersonen oder Gruppen von Personen. Unterschieden werden kann dabei zwischen verschiedener Frequenz, Übermittlungsart und Spezifität des Feedbacks: so erhalten Verbraucher beispielsweise zeitversetzt Berichte per Brief über den Energieverbrauch des vergangenen Monats, Quartals oder Jahres. Oft wird dabei der aktuelle Wert mit einem Vergleich des eigenen Verbrauchs in der Vergangenheit verbunden (historischer Vergleich). Alternativ finden kontinuierliche Rückmeldungen in Echtzeit statt, z.B. über einen Smart Meter (intelligente Zähler) oder In-Home-Displays (Bildschirme), die beispielsweise an der Heizung oder in der Dusche installiert sind. Außerdem kann sich das Feedback auf den allgemeinen Energieverbrauch oder den Verbrauch einzelner Geräte wie z.B. eines Haartrockners oder einer heißen Dusche beziehen.

Zahlreiche Studien verknüpfen Echtzeit-Verbrauchsfeedback per Smart Meter mit flexiblen Preisen (z.B. Buryk et al. 2015 und Allcott 2011a). Als Instrument der Nachfragesteuerung werden dabei üblicherweise zu Spitzenzeiten höhere Strompreise angesetzt, um Verbraucher zu Änderungen ihres Stromnutzungsverhaltens zu bewegen. Da sich die nicht-monetäre Intervention Feedback in diesen Fällen nicht von klassischen finanziellen Anreizen trennen lässt und die Intervention eher auf eine Verschiebung als eine Reduzierung des Energieverbrauchs abzielt, werden solche Interventionen in der vorliegenden Studie nicht berücksichtigt.

Sozialer Vergleich

Als Sozialer Vergleich wird die Bereitstellung von Informationen über den Stromverbrauch des Haushaltes im Verhältnis zu vergleichbaren Haushalten bezeichnet. Die Referenzgruppe sollte dabei ähnliche Eigenschaften aufweisen und beispielsweise denselben Energieversorger oder denselben Wohnort haben (Abrahamse et al. 2005). Ein Sozialer Vergleich ist somit eng mit Feedback zum eigenen Verbrauch verknüpft. Abhängig vom Referenzlevel kann der Vergleich des eigenen Verbrauchs beispielsweise mit dem durchschnittlichen Verbrauch aller Haushalte oder mit dem Verbrauch einer bestimmten Gruppe, etwa den effizientesten 10 Prozent aller Haushalte, erfolgen.

Ein Sozialer Vergleich wirkt *Biased Beliefs* über das eigene Verbrauchsverhalten im Vergleich zu anderen Haushalten entgegen. So kann beispielsweise eine Person, die sich selbst als umweltbewusst einstuft, ihren Energieverbrauch im Vergleich zu anderen systematisch unterschätzen. Eine solche Fehlwahrnehmung wird durch den Sozialen Vergleich korrigiert.

Drei Phänomene können dazu führen, dass infolge eines Sozialen Vergleiches Individuen ihren Energieverbrauch reduzieren: Zum einen zeigen viele Menschen bezugsabhängige Präferenzen (Kahneman 2003). Einen gängigen Referenzpunkt bilden soziale Normen. Vielen Menschen ist es wichtig Normen zu entsprechen, es steigert ihren Nutzen, während ein Abweichen zu negativem Nutzen führt (Schubert und Stadelmann 2015). Zum anderen nutzen Individuen bei Unsicherheit oft das Verhalten anderer als Orientierung, weil sie annehmen, dass diese über mehr Informationen – beispielsweise über das sozial Erwünschte – verfügen als sie selbst (Allcott und Mullainathan 2010; Delmas et al. 2013). Als Folge tendieren Leute dazu, ihre Handlungen an verbreitetes Verhalten anzupassen. Darüber hinaus wecken Soziale Vergleiche Wettbewerbsgedanken (Abrahamse et al. 2005). Dies ist besonders dann relevant, wenn der Verbrauch des Haushalts oberhalb des Durchschnitts oder oberhalb eines für den Haushalt gewünschten Grenzwerts liegt, wie beispielsweise noch nicht zu den effizientesten 10 Prozent zu gehören.

Selbstbindung

Unter Selbstbindung (oder Selbstverpflichtung) versteht man Interventionen, die ein Individuum dazu bringen, sich selbst auf zukünftige Handlungen zu verpflichten (Allcott und Mullainathan 2010). Beispiele hierfür sind mündliche oder schriftliche Zusagen oder das Versprechen, künftig Energie einzusparen (Abrahamse et al. 2005). Die Verpflichtung kann ein Versprechen gegenüber sich selbst sein oder öffentlich gemacht werden.

Psychologische Grundlage der freiwilligen Selbstbindung ist die Erkenntnis, dass manche Menschen zeitlich inkonsistente Präferenzen haben und sich darüber bewusst sind (O'Donoghue und Rabin 1999). So ist es nicht selten, dass Individuen eine kurzfristig Kosten verursachende Handlung wie beispielsweise Lernen oder die Recherche nach einem neuen Stromversorger so lange hinauszögern, dass sie am Ende überhaupt nicht stattfindet – und das, obwohl das Ergebnis der Handlung, neues Wissen oder eine günstigere Stromversorgung, vom Individuum immer noch gewünscht sind (vgl. bspw. O'Donoghue und Rabin 2008). Diese Verzerrung der eigenen Präferenzen wird durch die Gegenwartstendenz hervorgerufen, eine zu starke Gewichtung auf die Gegenwart, die sich in darin zeigt, Nutzenabschläge zwischen einem Genuss heute und einem Genuss morgen stärker zu gewichten als zwischen zwei zukünftigen Zeitpunkten (Laibson 1997). Dies kann in der Konsequenz zu zeitinkonsistentem Verhalten führen. Am Beispiel Energiesparen lässt sich der Sachverhalt folgendermaßen verdeutlichen: Ein Verbraucher nimmt sich zu Beginn der Woche für das folgende Wochenende vor, ein Thermostat für die Heizung zu kaufen, das ihm dabei helfen soll, Heizkosten zu sparen. Obwohl sich seine Präferenzen nicht geändert haben, geht er am Wochenende lieber einer Aktivität nach, die ihm sofortige Genugtuung verschafft. Er

entscheidet sich also am Wochenende gegen die Investition in langfristige Heizkosteneinsparungen, weil der Entscheidungszeitpunkt näher in die Gegenwart gerückt ist, plant aber weiterhin die Investition durchzuführen und prokrastiniert somit. Ein solches Verhalten ist vor allem aus Bereichen mit fernen Langzeitzielen wie beispielsweise Sparen bekannt (Thaler und Benartzi 2004).

Selbstbindungsverträge helfen dabei, solche zeitinkonsistente Präferenzen zu überwinden und das eigene Langzeitziel zu erreichen, indem sie als Motivationsquelle dienen: Sie stellen die gegenwärtige Situation in einen direkten Vergleich mit einem angestrebten zukünftigen Zustand (van Hoewelingen und van Raaij 1989). Liegt ein Versprechen gegenüber sich selbst vor, spricht ein Auseinanderklaffen zwischen gegenwärtiger Situation und angestrebtem Zustand eine persönliche Norm an: das Individuum will Erwartungen an sich selbst erfüllen. Ein öffentliches Versprechen erzeugt Erwartungen bei anderen und führt damit zu sozialem Druck (Abrahamse et al. 2005). Darüber hinaus fördert Selbstbindung aufgrund referenzabhängiger Präferenzen die Erreichung selbst gesteckter Ziele (Abrahamse et al. 2005): Verbraucher, die ein Versprechen gegeben haben, beurteilen ihre Leistung anhand des gesetzten Bezugspunkts. Erreichen sie das Ziel, erzeugt dies ein unmittelbares Erfolgsgefühl (und in ökonomischen Termini Nutzen), während ein Scheitern als negativ wahrgenommen wird.

Zielsetzung

Als Ergänzung zur Selbstbindung legen so genannte Zielsetzungen einen konkreten Referenzwert für das zukünftige Verhalten fest, zu dem man sich selbst verpflichtet. Anstatt einer generellen Zusage Energie zu sparen, verspricht ein Haushalt beispielsweise den Energiekonsum innerhalb des nächsten Monats um 10 Prozent zu reduzieren. Dieses Versprechen kann gegenüber sich selbst oder gegenüber einem externen Akteur wie dem Stromversorger oder der Verbraucherzentrale abgegeben werden. Dass neben der Festlegung eines Einsparniveaus auch eine Frist gesetzt wird, erleichtert die Bewertung der Bemühungen als Erfolg oder Scheitern. Dies erhöht den Druck, gleichzeitig aber auch die Motivation, weil die Zufriedenheit mit sich selbst oftmals von einer messbaren Leistung abhängt (van Houwelingen und van Raaij 1989). Der konkrete Zielwert kann dabei entweder von dem Haushalt selbst (als eine Form der Selbstbindung) oder extern (beispielsweise durch Institutionen) gesetzt werden. Selbst wenn das Ziel von außen vorgegeben und nicht selbst gesetzt wurde, vermag die Zielerreichung ein unmittelbares Erfolgs- oder Misserfolgsgefühl erzeugen (Bandura 1986). Ähnlich der Selbstbindung, versucht die Zielsetzung der zu starken Gegenwartspräferenz entgegenzuwirken.

Labeling

Ein Label ist ein Etikett, das Informationen über eine Ware auf leicht zugängliche Weise zusammenfasst. Dies kann durch eine Auswahl von Informationen zu den Produkteigenschaften auf dem Label oder durch eine grafische Visualisierung der wichtigsten Eckdaten erreicht werden. Im Bereich des Energieverbrauchs umfassen Labels beispielsweise Informationen über den jährlichen Energieverbrauch von Geräten, gemessen in Kilowattstunden, oder die Energieeffizienzstandards von Häusern.

Grundsätzlich gibt es zwei verschiedene Arten von Labels. Entweder handelt es sich um Gütesiegel, bei denen ein Gerät das Label als Auszeichnung erhält, wenn es bestimmte Standards erfüllt bzw. zu den energieeffizientesten seiner Produktklasse gehört. Die meisten dieser Gütesiegel sind freiwillige Kennzeichnungen. Ein Beispiel hierfür ist das US-amerikanische *Energy Star*-Label. Vergleichende Labels, die zweite Kategorie von Energielabels, bieten einen Gerätevergleich hinsichtlich ihrer Energieeffizienz, entweder kategorial (z.B. mit unterschiedlichen Balken)

oder mit einer kontinuierlichen linearen Skala. Diese Art der Zertifizierung ist in den meisten Ländern für alle Geräte einer Produktklasse obligatorisch, wie etwa beim EU-Energielabel.

Die Intervention Labeling adressiert mehrere der im Bereich der Energieeffizienz relevanten Internalitäten. Eine Möglichkeit, wie Verbraucher auf ein Label reagieren, ist, dass sie bestimmten Eigenschaften des Gutes mehr Aufmerksamkeit gewähren und dadurch systematische *Unachtsamkeit* in Bezug auf relevante Produktinformationen überwinden. Indem das Label die ausgewählten Informationen stärker hervorhebt, zielt es auf die Verfügbarkeitsheuristik ab. Damit wird ein kognitiver Prozess bezeichnet, der gut zugänglichen und leicht erinnerbaren Merkmalen einen stärkeren Einfluss auf Entscheidungen verleiht, während Informationen mit geringer Zugänglichkeit weitgehend ignoriert werden (Kahneman 2003). Über diesen Mechanismus trägt ein Label auch dazu bei, die so genannte *kostenintensive Informationsbeschaffung* zu reduzieren. Kosten sind hier zu verstehen im Sinne von Zeit, welche investiert werden muss, um die für die Kaufentscheidung relevante Informationen zu beschaffen. Diese Zeit hat Opportunitätskosten, da das Individuum alternativ etwas machen könnte, das ihm unmittelbaren Nutzen bringt (wie z.B. favorisierte Musik hören oder Sport treiben).

Die Wirkung der Verfügbarkeitsheuristik konnte jedoch auch unabhängig davon nachgewiesen werden, ob die Kosten für die Beschaffung relevanter Informationen, etwa über den durchschnittlichen Verbrauch oder die Lebenserwartung von Geräten, hoch oder niedrig sind (Schubert und Stadelmann 2015). Eine ähnliche Logik liegt dem sogenannten Konzentrations-Bias zugrunde: Individuen bevorzugen diejenigen Optionen, deren Nutzen sich konzentriert in der nahen Zukunft zeigt – und nicht erst im Laufe mehrerer Jahre, wie dies bei teuren, energieeffizienten Kühlschränken der Fall ist (siehe Köszegi und Szeidl 2012). Ein Label ermöglicht dem Verbraucher eine Antizipation des kumulierten Nutzens bereits zum Zeitpunkt der Kaufentscheidung, beispielsweise darüber, dass es die Energiekosten für die folgenden zehn Jahre zusammenfasst und mit denen eines weniger energieeffizienten Gerätes vergleicht.

Des Weiteren adressieren Labels systematische Fehleinschätzungen (*Biased Beliefs*) zu den Auswirkungen und dem Ausmaß verschiedener Energiesparmaßnahmen. Wenn beispielsweise ein Verbraucher das Einsparpotenzial eines Gerätes unterschätzt, investiert er – gemessen an seinen eigenen Präferenzen – möglicherweise zu wenig in Energieeffizienz (Allcott 2016). Ein Label, das die systematische Unterschätzung korrigiert, indem es beispielsweise die jährlichen Energiekosten mit denen eines weniger effizienten Gerätes vergleicht, reduziert diese Internalität und begünstigt damit energieeffizientere Kaufentscheidungen.

2.3. Vorausgehende wissenschaftliche Übersichtsarbeiten zur Thematik

Während es das Ziel des Analyseteils dieser Studie ist, den Stand der Forschung bezüglich der Wirksamkeit von verhaltensökonomischen Maßnahmen hinsichtlich Energieeinsparungen privater Haushalte zu ermitteln, sollen an dieser Stelle bereits existierende Überblicksarbeiten zu dem Thema kurz vorgestellt werden. Abrahamse et al. (2005) untersuchen in ihrem narrativen Systematic Review die Wirksamkeit ausgewählter Interventionen, welche auf eine Reduktion des Energieverbrauches privater Haushalte abzielen. Dazu zählen neben der verhaltensökonomischen Intervention Feedback auch die Bereitstellung von Informationen sowie Belohnungen. Die Autoren kommen zu dem Schluss, dass *Informationen* zwar einen Einfluss auf den Wissensstand der Haushalte haben, dieses Wissen jedoch nicht zwingend in Verhaltensänderungen oder Energieeinsparungen mündet. *Belohnungen* hingegen bewirken zwar Energieeinsparungen, sind jedoch eher von kurzfristiger Dauer. *Feedback*, insbesondere wenn es häufig gegeben wird, zeigt ebenfalls eine positive Wirkung auf energiesparendes Verhalten.

Mehrere Überblicksarbeiten haben speziell die Intervention Feedback untersucht (Roberts und Baker 2003; Darby 2006; Fischer 2008; Ehrhardt-Martinez et al. 2010). Die umfangreichste *Feedback*-Studie von Karlin et al. (2015) kommt anhand einer Meta-Analyse zu dem Schluss, dass Feedback zwar wirksam ist, jedoch mit signifikanten Schwankungen in den Ergebnissen. Delmas et al. (2013) untersuchen die Auswirkungen von *Informationen*, zu denen die Autoren auch *Feedback* zählen, auf den Energieverbrauch und finden im Durchschnitt einen erheblichen Reduktionseffekt. Lokhorst et al. (2013) arbeiten in einer Meta-Analyse zu Selbstbindungsinterventionen im Umweltbereich basierend auf 19 Studien heraus, dass diese kurz- und langfristig umweltfreundliches Verhalten fördern. Sie weisen jedoch darauf hin, dass dies besonders dann der Fall ist, wenn Selbstbindung mit anderen Interventionen gekoppelt ist. Darüber hinaus identifizieren sie als Schwachstelle, dass die konkreten Mechanismen, die dieser Wirkung zugrunde liegen, bislang nicht klar identifiziert werden konnten.

Delmas et al. (2013) wie auch Abrahamse et al. (2005) kommen außerdem zu dem Schluss, dass Studien mit besseren Evaluationsdesigns tendenziell eine geringere Wirksamkeit der Interventionen dokumentieren. Dies zeigt mögliche methodische Probleme in der betrachteten Literatur auf. Mit Ausnahme von Andor und Fels (2017a) hat keiner der vorliegenden Überblicksartikel als Auswahlkriterium, ob in den betrachteten Studien eine Methode verwendet wird, mit der kausale Effekte der Interventionen identifiziert werden können.

Die eigenen Vorarbeiten aus Andor und Fels (2017a) erweitern wir in dieser Studie insbesondere durch die Anwendung meta-analytischer Methoden. Diese erlauben nicht nur eine qualitative, sondern auch eine quantitative Auswertung der bestehenden empirischen Literatur. Zudem ist die Suche um eine zusätzliche Datenbank erweitert worden und die weit verbreitete Intervention Feedback wird mit einbezogen.²

Im Vergleich zur weiteren bisherigen Forschung unterscheidet sich die vorliegende Studie demnach dahingehend, dass

- in ihr die aktuelle Forschung abgedeckt wird – die jüngsten vorausgehenden Studien von Karlin et al. (2015) und Lokhorst et al. (2013) umfassen zum Beispiel nur Studien bis 2010, und somit einen Gutteil der in diese Meta-Analyse aufgenommenen Studien nicht (s. Kapitel 6.2);
- ein breiteres Spektrum an verhaltensökonomischen Maßnahmen untersucht wird;
- durch die Anwendung meta-analytischer Methoden eine quantitative Aggregation der Einzeleffekte aus der Primärliteratur ermöglicht wird, wie es bislang einzig Karlin et al. (2015) und Lokhorst et al. (2013) angestellt haben;³
- ausschließlich Studien betrachtet werden, die das Potenzial haben, kausale Effekte zwischen der Intervention und dem Energieverbrauch zu identifizieren.

² In Andor und Fels (2017b) war diese Intervention zwar enthalten, allerdings wurde die Suche in der vorliegenden Studie substantiell erweitert (bspw. bei dem berücksichtigten Zeitrahmen als auch den Suchstrategien).

³ Delmas et al. (2013) nutzen in ihrer ansonsten sehr aufschlussreichen Studie zwar auch meta-analytische Methoden, nutzen dabei jedoch einen nicht anerkannten Ansatz, bei dem nicht-standardisierte *t*-Statistiken als Effektgrößen dienen (Becker und Wu 2007; Campbell Collaboration 2014).

3. Methoden

3.1. Kriterien für die Einbeziehung von Studien in die Meta-Analyse

Der folgende Kriterienkatalog basiert auf dem gängigen „PICOS“-Schema, benannt nach dem englischen Akronym für Participants, Interventions, Comparisons, Outcome measures, und Study design (Campbell Collaboration 2014).

Studienobjekte

In die Analyse werden alle Studien einbezogen, die als Studienobjekte Haushalte oder Einzelpersonen in Haushalten untersuchen. Hierzu zählen beispielsweise auch Räume in Studentenwohnheimen. Ausgeschlossen werden hingegen Studien, die sich ausschließlich auf Firmen oder öffentliche Institutionen konzentrieren. Als Zielländer der Interventionen werden nur Studien berücksichtigt, die in Industrieländern durchgeführt wurden. Dazu zählen alle Länder, die von der Weltbank als „Länder mit hohem Einkommen“ klassifiziert wurden (siehe World Bank 2017). Zu den Ländern, die aufgrund dieses Kriteriums ausgeschlossen wurden, zählen u.a. China, Kolumbien und Russland.

Interventionen

Alle Studien, welche die Effekte der fünf im Fokus stehenden Interventionen – Feedback, Sozialer Vergleich, Selbstbindung, Zielsetzung und Labeling – untersuchen, werden in die Meta-Analyse einbezogen. Hierzu zählen auch Studien, in denen Kombinationen der genannten Interventionen untereinander oder mit anderen Interventionstypen analysiert werden.

Vergleichsintervention

Die Vergleichsbedingungen, die eine Studie verwenden muss, um sich für eine Aufnahme in unsere Meta-Analyse zu qualifizieren, entsprechen dem üblichen Vorgehen in der Literatur. Sie sind: (i) keine Intervention, (ii) gestaffelte Umsetzung oder (iii) „business as usual“. Damit werden Studien nicht einbezogen, die einen Vergleich der Studienobjekte mit sich selbst in einem Zeitraum vor der Intervention dokumentieren (vorher-nachher), selbst wenn für anderweitig beeinflussende Faktoren statistisch kontrolliert wurde.

Ergebnisvariablen

Unsere Meta-Analyse berücksichtigt alle Studien, die als Ergebnisgröße den Effekt auf den Energieverbrauch von privaten Haushalten oder Einzelpersonen dokumentieren. Energieverbrauch umfasst dabei die Ressourcen Gas, Strom und Wasser. Für den Bereich Labeling werden auch solche Studien einbezogen, die den Energieverbrauch indirekt betreffen, also als Ergebnisvariable beispielsweise die Zahlungsbereitschaft für energieeffiziente Geräte untersuchen.

Für die anderen Interventionen werden solche Studien, die bloß indirekte Effekte auf den Energiekonsum betrachten, ausgeschlossen. Beispiele hierfür sind Vellei et al. (2016) und Hille (2016). Vellei et al. (2016) untersuchen inwiefern Feedback durch eine App die Raumtemperatur und die Kleiderwahl von Studenten in Wohnheimen beeinflusst. Hille (2016) ermitteln den Effekt verschiedener Rechnungsvarianten (pauschal und verbrauchsabhängig) auf das Warmwassersparverhalten und die Raumtemperatur.

Forschungsdesign

Um kausale Effekte identifizieren und vergleichen zu können, bezieht die Meta-Analyse nur Studien mit ein, welche die Wirkung der Interventionen mit einem Studiendesign untersucht haben,

deren Ergebnisse grundsätzlich als intern valide und damit glaubwürdig eingestuft werden können (s.a. Greenhalgh 2014). Dies umfasst insbesondere folgende rigorose Methoden der Kausalanalyse: (i) randomisierte, kontrollierte Feldexperimente (RCT), (ii) Matching, (iii) Regressions-Diskontinuitäts-Analyse (RDD), (iv) Differenz-von-Differenzen (Diff-in-Diff), (v) Fixe Effekte, (vi) Instrumentvariablen-Ansätze (IV) und (vii) natürliche Experimente. Ausführlich werden diese Methoden u.a. in Angrist und Pischke (2009) sowie Imbens und Wooldridge (2009) dargestellt, sowie in deutscher Sprache in Bauer et al. (2009) und Boockmann et al. (2014). Als absolutes Minimum wird eine Stichprobengröße von mindestens 20 unabhängigen Beobachtungen vorausgesetzt.

Weitere Kriterien

Wir beziehen alle Studien ein, die in referierten Zeitschriften oder anerkannten Working-Paper-Serien veröffentlicht wurden (siehe Abschnitt 3.3). Bei keiner Suche wurde der Zeitraum der Publikation begrenzt.

3.2. Systematische Suche relevanter Primärliteratur

In einem ersten Schritt erfolgte eine systematische Datenbanksuche. Diese erstreckte sich auf die drei Datenbanken JSTOR, ScienceDirect, und EconLit, somit die jeweils relevanteste generelle, verlagsbetriebene bzw. ökonomiespezifische Datenbank. Sie wurden anhand vorab festgelegter, identischer Schlagwort-Kombinationen durchsucht (s. Anhang A.1). In den Datenbanken ScienceDirect und JSTOR wurden neben ökonomischen Studien auch solche aus den Disziplinen Psychologie, Sozialwissenschaften, Umweltwissenschaften und Energie gesucht, um dem interdisziplinären Charakter der Forschung zu diesem Thema Rechnung zu tragen.

Zusätzlich zu den Datenbankrecherchen haben wir eine Rückwärtssuche angewandt, indem wir die Referenzlisten der vorliegenden Überblicksstudien von Abrahamse et al. (2005), Delmas et al. (2013), Karlin et al. (2015) und Lokhorst et al. (2013) sowie der überblickartigen Studien zu Labeling von Rohling und Schubert (2013) und Schubert und Stadelmann (2015) vollständig aufgenommen und die einzelnen Einträge auf die in der vorliegenden Meta-Analyse relevanten Kriterien überprüft haben.

Alle Artikel, die über die beschriebene Suche identifiziert werden konnten, wurden in das Literaturmanagementprogramm Citavi exportiert. Nach einer Dublettenbereinigung, die auch das Aussortieren von Arbeitspapieren (Working Papers) umfasste, welche in einer späteren Version als Journal-Artikel einbezogen wurden, wurden in einem zweiten Schritt die Abstracts sämtlicher Artikel von wissenschaftlichen Hilfskräften unter Anleitung der Autoren dieses Berichts gelesen und anhand eines standardisierten Bewertungsbogens kodiert. Dieser Bewertungsbogen umfasste die sieben in Tabelle 3 aufgeführten Bewertungsfragen, die sich an den oben vorgestellten PICOS-Kriterien orientieren. In Zweifelsfällen wurde zusätzlich zum Abstract der Volltext des Artikels als Entscheidungsgrundlage mit einbezogen. Falls Zweifel bestehen blieben, sowie grundsätzlich bei der anspruchsvollsten Bewertungsfrage 7 zur Methode der Kausalanalyse, wurde die Entscheidung über die Erfüllung des jeweiligen Kriteriums durch die Autoren dieses Berichts getroffen.

Tabelle 3
Bewertungsfragen als Grundlage für die Einbeziehung der Primärliteratur

Auswahlkriterium	Bewertungsfrage
Primärliteratur	1. Handelt es sich bei der Studie um Primärforschung?
Studienobjekte	2. Studie wurde in einem Industrieland durchgeführt?
	3. Studie bezieht sich auf private Haushalte oder Einzelpersonen in privaten Haushalten als Studienobjekte?
Interventionen	4. Studie bezieht sich auf mindestens eine der ausgewählten Interventionen (Feedback, Sozialer Vergleich, Selbstbindung, Zielsetzung, Labeling)?
Ergebnisvariablen	5. Studie bezieht sich auf mindestens eine der ausgewählten Ergebnisvariablen (Strom, Gas, Wasser), für Labeling auch mittels indirekter Ergebnisvariablen?
Forschungsdesign	6. Bei der Studie handelt es sich um angewandte Forschung?
	7. Studie wendet Methode der Kausalanalyse an?

Die identifizierten Artikel können in sechs verschiedene Kategorien aufgeteilt werden, für die Tabelle 4 beispielhaft Artikel auflistet. Diese Meta-Analyse berücksichtigt nur Artikel der Kategorie 6. Artikel der Kategorie 1 dienen zusätzlich als Hintergrundmaterial.

Tabelle 4
Übersicht über Kategorien und beispielhafte Artikel

Kategorie	Relevante Bewertungsfrage*	Art der Artikel	Beispielhafte Artikel
1	1	Relevante Meta-Analysen oder Systematic Reviews	Delmas et al. 2013; Abrahamse et al. 2005
2	2, 3 oder 5	Empirische Studien, die im Bereich Energieverbrauch angesiedelt sind, sich aber entweder nicht auf die drei ausgewählten Ergebnisvariablen oder nicht auf den Energieverbrauch von privaten Haushalten in Industrieländern beziehen	Bedwell et al. 2014; Dogan et al. 2014; Dowd et al. 2012; Litvine und Wustenhagen 2011; Murtagh 2013; Pullinger et al. 2014; Stillwater und Kurani 2013; Yeomans und Herberich 2014
3	4	Empirische Studien, die im Bereich Energieverbrauch angesiedelt sind, aber nicht konkret eine der ausgewählten fünf Interventionen testen	Aste et al. 2013; Brutscher 2011; Di Cosmo et al. 2014; Lang und Okwelum 2015; Richalet et al. 2001; Wang et al. 2011
4	6	Theoretische Modellierungen	Bauwens et al. 1994; Darby 2006; Lutzenhiser 1992; Shen et al. 2015; Steg 2008; Zhou et al. 2011
5	7	Interventionsstudien oder Surveys, die empirische Daten zu den ausgewählten Interventionen erheben, aber keine Kausalanalyse ermöglichen	Buchanan et al. 2014; Hargreaves et al. 2010; Hargreaves et al. 2013; Heberlein und Warringer 1983; Ivanov et al. 2013; Midden et al. 1983; Mills und Schleich 2010; Oltra et al. 2013; Oseni et al. 2013; Paetz et al. 2012
6	-	Interventionsstudien oder Surveys, die über quantitative Daten eine Kausalanalyse zur Fragestellung ermöglichen	s. Kapitel 6.2

* siehe Tabelle 3

3.3. Datensammlung

Der finale Schritt, das Lesen und Exzerpieren der ausgewählten Artikel, wurde erneut von wissenschaftlichen Hilfskräften unter Anleitung der Autoren dieses Berichts durchgeführt. Für die Exzerpte wurden die Volltexte aller einbezogenen Artikel gelesen und die notwendigen Informationen herausgefiltert. Der daraus resultierende Datensatz enthält vor allem Informationen zu folgenden Aspekten:

- Allgemeine Studieninformationen (z.B. Autor, Typ der Veröffentlichung)
- Studiendesign (z.B. Interventionstyp, mögliche Kombinationen mit anderen Interventionstypen, Design der Intervention, Methode der Kausalanalyse, Dauer der Intervention)
- Datensatz (die Anzahl der in den Studien betrachteten Haushalte bzw. Individuen)
- Abhängige Variable (z.B. Typ (Strom, Gas, Wasser), Maßeinheit (z.B. kWh/Monat), Mittelwert)
- Effektschätzungen (z.B. Effektgröße, Signifikanzniveau).

Eine vollständige Liste der Variablen kann Anhang A.2 entnommen werden. Die Daten dienen vor allem zur Berechnung der Effektgröße (siehe Kapitel 3.4), zur korrekten Identifikation und Einordnung der Studien sowie für mögliche Moderatoranalysen. Diese Analysen dienen dazu Variablen (= Moderatoren) zu identifizieren, die die Größe von Effekten sowie deren Variation über Studien zu erklären vermögen – wie zum Beispiel die Dauer der jeweiligen Intervention. Auch in diesem Exzerpierungsschritt wurden die Autoren dieses Berichts eingeschaltet, um Unklarheiten und Unstimmigkeiten in den Exzerpten auszuräumen und um sämtliche Angaben zu Kerndaten der Studien wie den extrahierten quantitativen Effekten zu prüfen.

3.4. Effektgrößenberechnung

Die quantitative Meta-Analyse, deren Ergebnisse in Kapitel 4 vorgestellt werden, bündelt Einzelergebnisse verschiedener Studien zu Punktschätzern des allgemeinen Effekts der untersuchten Maßnahmen. Zusammen mit den zugehörigen Konfidenzintervallen liefern diese Punktschätzer eine – im Vergleich zu einzelnen Studien – erhöhte statistische Aussagekraft sowie die wahrscheinliche Bandbreite des jeweiligen „wahren“ Effektes (Cohn und Becker 2003). Im Einklang mit unserer Suchstrategie werden die Daten getrennt über zwei Dimensionen hinweg „gepoolt“, also zusammengefasst. Diese beiden Dimensionen sind der Interventionstyp und der Typ der Ergebnisvariable (siehe Tabelle 5). Kombinierte Interventionen (z.B. Feedback + Sozialer Vergleich) und aggregierte Ergebnisvariablen (v.a. Strom + Gas) werden ebenfalls eigenständig betrachtet und somit separat zusammengefasst.

Tabelle 5

Zur Zusammenfassung von Effektgrößen herangezogene Dimensionen

Interventionen	Ergebnisvariablen
1 Feedback	1 Strom
2 Sozialer Vergleich	(bei Labeln auch indirekt)
3 Selbstbindung und Zielsetzung	2 Gas
4 Labeling	3 Wasser
& Kombinationen dieser Interventionen	& Energie aggregiert

Um valide Ergebnisse zu erzielen, muss in einer Meta-Analyse darauf geachtet werden, dass jede Studie nur mit einem einzigen Effektschätzer in die Analysen der einzelnen Interventions-Ergebnisvariablen-Kombinationen eingeht. So wird gewährleistet, dass keine Studie unangemessenes Übergewicht bei der Berechnung des allgemeinen Effekts erhält und dass die statistische Unabhängigkeit der Daten gewährleistet ist (Borenstein et al. 2009a). Dies bedeutet, dass bei Vorhandensein mehrerer abhängiger Effektgrößen innerhalb einer Studie die entsprechenden Schätzungen hierarchisiert oder kombiniert werden müssen. Zu beachten ist hierbei, dass wir verschiedene Treatmentgruppen desselben Interventionstyps innerhalb eines Experimentes als unabhängige Gruppen werten und somit getrennt in die Berechnung des allgemeinen Effektes aufnehmen.

Diese Hierarchisierung oder Kombination (auch als „Synthetisierung“ von Effektgrößen bezeichnet) war in etwas mehr als der Hälfte der Studien nötig. Costa und Kahn (2013) beispielsweise analysieren den Effekt einer Maßnahme mit Sozialem Vergleich auf den Energieverbrauch (selbe Intervention und selbe Ergebnisvariable) anhand verschiedener Spezifikationen. In diesen Fällen wurden zunächst die in Anhang A.3 aufgeführten Hierarchisierungs- und Auswahlkriterien systematisch herangezogen. So wurden beispielsweise solche Ergebnisse bevorzugt, die die gesamte Studienpopulation anstatt einer Untergruppe der Studienpopulation betrachten. Eine klare Präferenz für bestimmte Effektschätzungen konnte jedoch nicht immer bestimmt werden. In diesen Fällen mussten die verschiedenen Effektgrößen, wie in Anhang A.4 beschrieben, synthetisiert werden.

Um grundsätzlich Vergleichbarkeit zwischen Schätzergebnissen von Studien herzustellen, die auf unterschiedlichen Methoden und Stichprobengrößen basieren, ist die Berechnung einer standardisierten Effektgrößenstatistik notwendig. Während auch andere Ansätze Verwendung finden, ist bei der Ermittlung von Maßnahmeneffekten zu empfehlen, entweder die standardisierte Mittelwertsdifferenz (*standardised mean difference, SMD*) oder den natürlichen Logarithmus der response ratio (*InRR*) zu verwenden (Nakagawa et al. 2017; Hedges et al. 1999). *InRR* eignet sich in unserem Fall insbesondere aus dem Grund, dass viele der betrachteten Studien den Effekt lediglich als prozentuale Veränderung berichten, wodurch die Berechnung des *SMD* häufig unmöglich ist. Zudem sind die Werte der response ratio einfacher interpretierbar.⁴ In Kapitel 4 werden wir daher primär *InRR*-Ergebnisse darstellen. Neben einem Punktschätzer wird der Standardfehler des Schätzers berechnet, um die oben erwähnte wahrscheinliche Bandbreite dieses Schätzers zu ermitteln – technisch anhand von 95%-Konfidenzintervallen. Die verwendeten Formeln sind Anhang A.4 zu entnehmen.

Um die so berechneten einzelnen Effektgrößenstatistiken in der Gesamtschau meta-analytisch zu analysieren, wenden wir die populäre Inverse-Varianz-Gewichtung mit zufälligen Effekten (*inverse-variance random effects*) an, bei der die „*between-study variance*“ mit der Methode von DerSimonian und Laird (1986) berechnet wird. Die Wahl einer Modellierung mit zufälligen statt fixen Effekten erscheint angemessen aufgrund der vermutlich ausgeprägten kontextuellen Variation zwischen den Studien (z.B. im konkreten Interventionsdesign und Implementierungsprozess), und dem generellen Vorteil, dass die Modellierung eine bessere Generalisierbarkeit auf

⁴ Ferner sind die Voraussetzungen dafür erfüllt, *InRR* sinnvoll anzuwenden. Insbesondere handelt es sich bei den Hauptergebnisvariablen Strom-, Gas- und Wasserkonsum um Verhältnisskalendaten, die (i) mittels einer metrischen Skala gemessen werden (z.B. kWh) und (ii) nur in seltenen Fällen den Wert null annehmen, der gleichzeitig das Minimum darstellt (Borenstein et al. 2009b, Houle et al. 2011).

Kontexte erlaubt, die nicht von den betrachteten Studien abgedeckt sind (Hunter und Schmidt 2000).⁵

Darüber hinaus führen wir (Weighted Least Squares (WLS)-basierte) Metaregressionsanalysen durch, die meta-analytische Instrumente mit dem Regressionsansatz kombinieren, um verschiedene Einflussfaktoren auf Effektgrößen besser bewerten zu können. Die Effektgrößen werden mit der Software Stata (Stata Corporation, College Station, TX, USA) mittels der Befehle *metan* und *metareg* geschätzt. Die Ergebnisse werden entsprechend der in der Literatur üblichen Gepflogenheiten dargestellt (vgl. Stanley et al. 2013).

⁵ Während die Methode von DerSimonian und Laird seine Schwachstellen hat (Veroniki et al. 2016), vermögen alternative Ansätze noch nicht zu überzeugen (Partlett und Riley 2017) und – wie Kelley und Kelly (2012) argumentieren – sind die Unterschiede zwischen bestehenden Methoden gering.

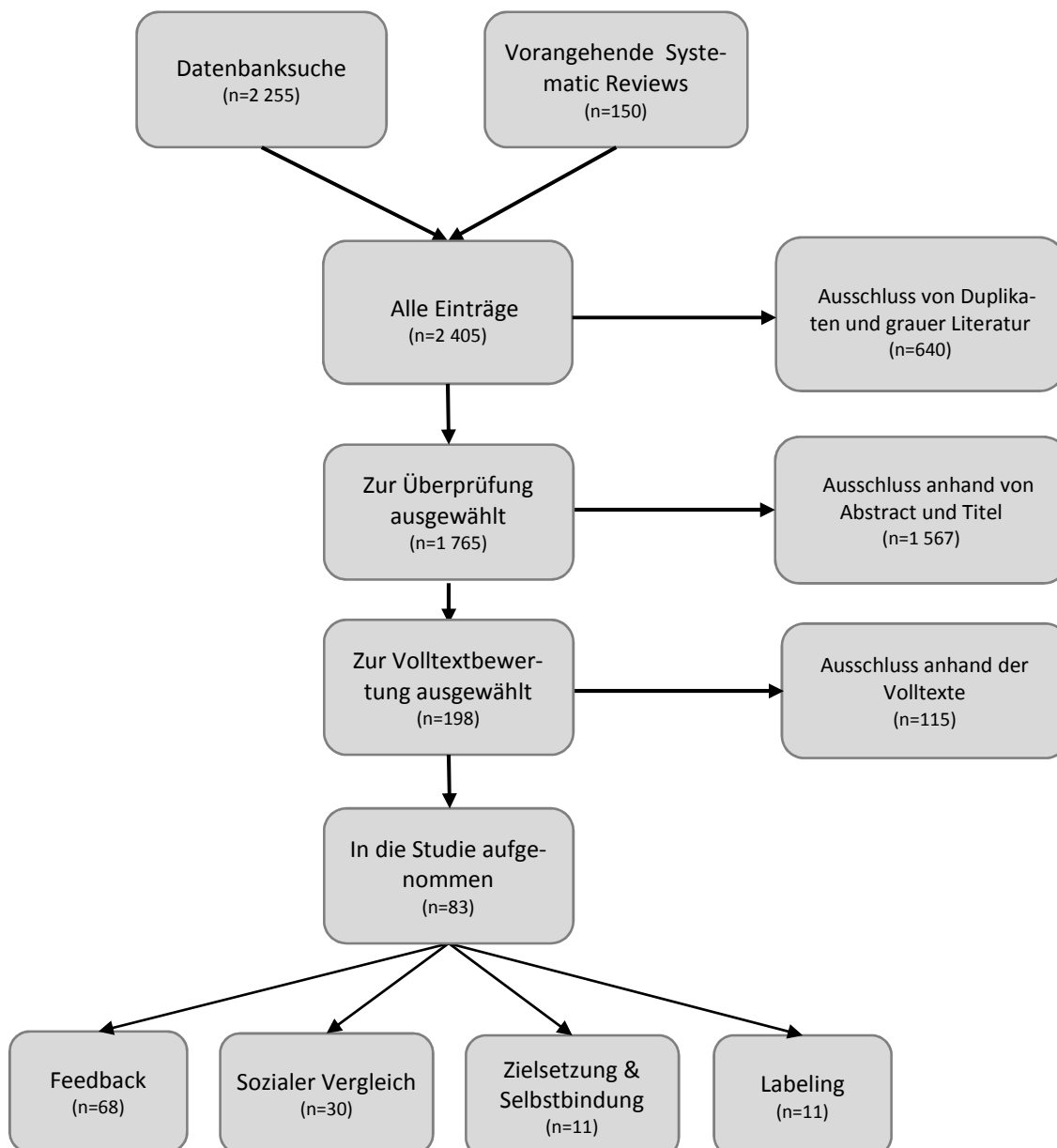
4. Resultate

4.1. Ergebnisse der systematischen Suche

Die in Kapitel 3 beschriebene Datenbanksuche erzielte 2 255 Treffer (siehe Abbildung 1). Zusätzlich umfassten vorangehende Übersichtsarbeiten 150 Artikel, die als potenziell relevant identifiziert wurden. Von diesen wurden 640 als Duplikate erkannt und aussortiert. Von den restlichen 1 765 Artikeln hielten in einem ersten Schritt (der Durchsicht von Abstract und Titel) 198 den Kriterien aus Tabelle 3 stand. Nach der Prüfung der Volltexte verblieben 83 Artikel, die allen Kriterien entsprechen und in die Analyse einbezogen wurden. Studien, die mehr als eine Intervention untersuchen, werden in der untersten Zeile der Abbildung mehrfach gezählt.

Abbildung 1

Überblick über den Such- und Selektionsprozess



Anstatt auf Artikel beziehen wir uns im Folgenden auf eigenständige Studien. Tabelle 6 zeigt in welchem Verhältnis die einbezogenen Studien und Artikel zueinander stehen. Während der Großteil der Artikel die Ergebnisse einer Studie vorstellt, gibt es zwei Studien, deren Kurz- und Langfristeffekte auf zwei unterschiedliche Artikel verteilt besprochen wurden, sowie einige Artikel, die mehrere Studien beinhalten. Dabei handelt es sich um unterschiedliche Interventionen mit unterschiedlichen Populationen, oder unterschiedliche Populationen, die der gleichen Interventionen ausgesetzt wurden, oder ein und dieselbe Population, die zu unterschiedlichen Zeitpunkten unterschiedlichen Interventionen erhalten. Darüber hinaus gibt es unter anderem fünf Studien, die zwar alle im Rahmen von Kooperationen zwischen dem Dienstleister Opower und Energieversorgern in den USA generiert wurden, sich jedoch auf Populationen beziehen, die als statistisch unabhängig gelten können.⁶

Gemäß dieser Definition sind 95 Studien im Rahmen dieser Suche gefunden worden.

Tabelle 6

Artikel-Studien-Kombinationen und beispielhafte Artikel

Artikel-Studien-Kombinationen	Beispielhafte Artikel	Anzahl Studien
Mehrere Artikel, eine Studie	Ferraro et al. (2011) & Ferraro und Price (2013)	1
Ein Artikel, eine Studie	Tiefenbeck et al. (2013), Allcott und Rogers (2014)	70
Ein Artikel, mehrere Studien	Houde (2014), Dolan und Metcalfe (2015)	24
		95

4.2. Finale Auswahl von Effektschätzern und Studien

Aus den 95 Studien wurden zunächst insgesamt 638 Effektschätzer in diese Meta-Analyse aufgenommen, da sie grundsätzlich alle in Kapitel 3.1 erwähnten Kriterien erfüllen. Dennoch müssen in einem letzten Schritt vor der eigentlichen Analyse noch bestimmte Schätzer und damit auch komplette Studien ausgeschlossen werden. Ausschlussgründe sind in Tabelle 7 aufgeführt.

Hauptgrund ist die in Kapitel 3.4 erwähnte Hierarchisierung der Effektschätzer, weswegen insgesamt 303 Schätzer ausgeschlossen wurden. Entsprechend der Logik der Hierarchisierung wurden hierbei keine kompletten Studien ausgeschlossen. Dies war hingegen der Fall bei sechs Studien, die zwar prinzipiell von den Autoren als randomisierte, kontrollierte Feldexperimente beschrieben werden, die Kontrollgruppe jedoch nicht in ihre Analyse aufnahmen. Stattdessen wurden lediglich einfache Vorher-Nachher-Vergleiche für die Treatmentgruppe angestellt. Aufgrund genereller zeitlicher Trends sind die so bestimmten Effekte in der Regel verzerrt und lassen keine Differenzierung zwischen zeitlichem Trend und eigentlichem Effekt der Intervention zu – dies lässt sich sehr gut an Fielding et al. (2013) veranschaulichen, die zeitliche Trends des Wasserkonsums untersuchen. Da sie solche Trends tatsächlich finden, können sie die mit den untersuchten Interventionen einhergehenden Konsumveränderungen nicht analysieren. Die Studie kann daher

⁶ Seit März 2008 hat Opower sogenannte Home Energy Reports an die Kunden verschiedener kooperierender Energieversorger gesendet (s.a. Kapitel 4.4.2). Jedoch nutzt von den fünf einbezogenen Studien nur Allcott (2011b) Daten von allen Projekten aus den Jahren 2008 und 2009. Ayres et al. (2013) untersuchen den Energiekonsum von Haushalten aus dem Sacramento County in Kalifornien und dem Puget Sound in Washington. Auch Costa und Kahn (2013) bestimmen den Einfluss solcher Reports auf der Grundlage von Daten eines einzelnen Energieversorgers aus dem Westen der Vereinigten Staaten. Die fünfte Studie untersucht Haushalte aus drei verschiedenen Regionen (Allcott und Rogers 2014).

keine kausalen Effekte der Interventionen messen und findet sich ebenfalls in der Tabelle unter den ausgeschlossenen Studien.

Tabelle 7

Weitere Ausschlüsse von Studien und Effektschätzern

Ausschlussgrund	Anzahl Ausschlüsse		Ausgeschlossene Studien
	Schätzer	Studien	
RCT ausschließlich mit Vorher-Nachher-Vergleichen	38	6	Bekker et al. (2010) ; Bittle et al. (1979a); Grønhoj und Thøgersen (2011); Kua und Wong (2012); Schultz et al. (2007); van Houwelingen und van Raaij (1989)
Quasi-experimentelle Studien ausschließlich mit Vorher-Nachher-Vergleichen	19	1	Staats et al. (2004)
Nicht-bevorzugte Effektschätzer	303	-	-
Effektschätzer nur qualitativ genutzt (da nur Trendanalyse des Verbrauchs)	6	1	Fielding et al. (2013)
Verbleibende Schätzer und Studien	272	87	

Entsprechend der untersten Zeile von Tabelle 7, gehen insgesamt 272 Schätzer aus 87 Studien in die weiteren Untersuchungen ein. Unter diesen Schätzern sind alleine 80, die den Effekt der jeweiligen Intervention ausschließlich in Prozent angeben, weswegen die Berechnung der standardisierten Mittelwertsdifferenz *SMD* für diese Schätzer nicht möglich ist.⁷ Entsprechend fokussiert sich die Darstellung der Ergebnisse auf die alternative Effektgrößenstatistik, *InRR*. Da sich die Schätzer in wenigen Fällen auf die gleiche Kombination aus Intervention und Ergebnisvariable beziehen, mussten sie entsprechend der Beschreibung in Kapitel 3.4 zusammengefasst (gepoolt) werden.

4.3. Studienbeschreibung

Tabelle 8 fasst die Hauptcharakteristika der betrachteten Studien zusammen. Das Gros der Studien wurde 2010 oder später veröffentlicht, wobei eine beachtliche Anzahl an Studien sich bereits in den 1970er und 1980er mit den betrachteten verhaltensökonomischen Interventionen (mit Ausnahme von Labeling) befasst haben. Zudem wurde die Mehrheit in referierten Zeitschriften publiziert. Die Mehrzahl der in dieser Meta-Analyse berücksichtigten Studien wurde in den Vereinigten Staaten durchgeführt. Etwa halb so viele Studien fanden in verschiedenen europäischen Ländern statt, die meisten davon in den Niederlanden und dem Vereinigten Königreich, einige aber auch in skandinavischen Ländern und Deutschland. Wenige weitere Studien stammen aus dem asiatischen und ozeanischen Gebiet. Lediglich für die Intervention Labeling überwiegt die Anzahl der Studien aus Europa. Die Intervention Zielsetzung bzw. Selbstbindung wurde lediglich in den USA, in den Niederlanden, im Vereinigten Königreich und in Japan durchgeführt.

⁷Zudem ist zu beachten, dass diese Prozentangabe z.T. unterschiedlich konstruiert sind: z.B. unterscheiden sie sich dadurch, dass für ausgewählte Faktoren kontrolliert wird oder nicht, dass sie die Veränderung auf den über die Zeit aggregierten Verbrauch oder auf den Verbrauch an zwei bestimmten Zeitpunkte beziehen, und dass Unterschiede in Prozentpunkten und nicht in Prozenten gemessen werden (wobei letzteres insbesondere in den experimentellen Studien kein Problem darstellen sollte, da sich die Verbräuche in der Baseline i.d.R. nicht unterscheiden). Auf diese Unterschiede wird im Folgenden nicht weiter eingegangen.

Tabelle 8
Zusammenfassung der betrachteten Studien

Publikationsdatum	Art der Publikation		Untersuchungsregion		
Vor 1980	8	Zeitschriftenartikel	79	Nordamerika	53
1980 – 1989	11	Working-Paper	10	Europa	28
1990 – 1999	6			Asien und Ozeanien	6
2000 – 2009	8				
2010 – 2014	35				
2015 – 2017	19				
Gesamt	87		Gesamt 87	Gesamt	87

Die Anzahl an Studienteilnehmern variiert stark (Abbildung 2). Einigen Studien liegen nur Daten von weniger als 50 Teilnehmern zugrunde, anderen wiederum von über 10 000.

Abbildung 2
Anzahl der Studienteilnehmer

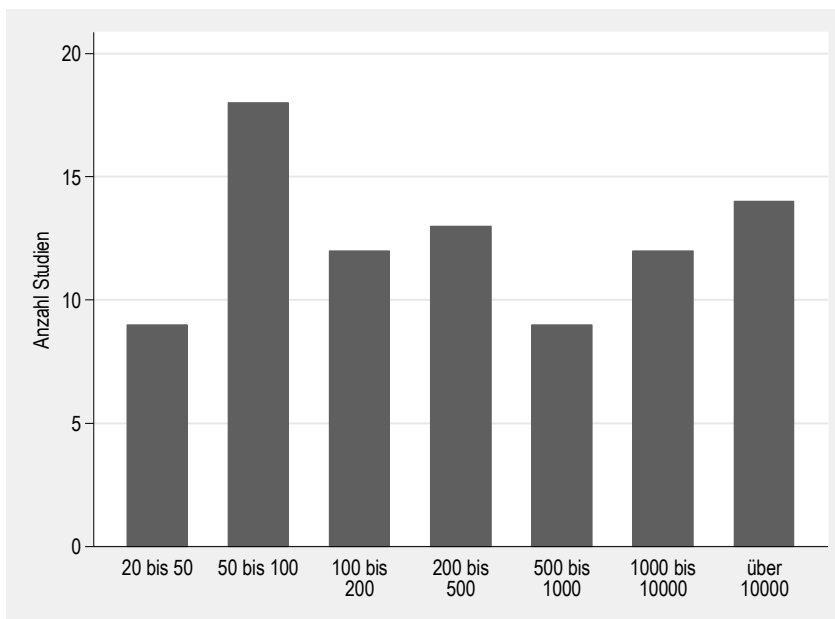


Tabelle 9 führt weitere Charakteristika der aufgenommenen Studien auf. Die den Studien zugrundeliegenden Daten wurden in knapp der Hälfte der Fälle von den Autoren selbst erhoben. Ähnlich viele Studien nutzen Daten privater Unternehmen oder von Nichtregierungsorganisationen (NRO), in den meisten Fällen von Energieunternehmen. Der weit überwiegende Anteil der Studien analysiert Daten, die „im Feld“ erhoben wurden. Lediglich neun Studien wurden im Labor oder über Online-Plattformen („Anderer Erhebungskontext“) durchgeführt, wobei diese Studien mehrheitlich die Intervention Labeling erforschten. In über der Hälfte der Studien wurden die genutzten Daten für Forschungszwecke erhoben, meist speziell für die verhaltensökonomische Forschung. Alternativ oder ergänzend wurden automatisch generierte Daten verwendet, wozu insbesondere von Energieunternehmen zur Verfügung gestellte Verbrauchsdaten zählen. Fast alle Studien wurden auf Haushaltsebene durchgeführt. Lediglich Studien, die die Intervention

Labeling untersuchen, untersuchen deren Effekt auch auf Individualebene. Dies hängt damit zusammen, dass die Studien zum Teil lediglich hypothetische Entscheidungen im Labor abfragen (siehe dazu *Ergebnisvariablen* weiter unten).

Tabelle 9
Eigenschaften der Datenerhebung der Studien

Datenerheber		Erhebungskontext		Erhebungsanlass		Untersuchungseinheit	
Autoren	45	Feld	78	Nudging-Forschung	44	Haushalt	74
Privates Unternehmen/ NRO	40	Labor	5	Automatisch generiert	45	Individuum	15
Öffentlicher Träger	8	Anderer	4	Für andere Forschung	4		
Teilnehmer selbst	1						
Gesamt	94	Gesamt	87	Gesamt	93		89

Mehrfachnennungen möglich

In Tabelle 10 werden die Unterschiede zwischen den Studien in der Art des Forschungsdesigns und der Berechnungsmethode dargestellt. Bis auf wenige Ausnahmen sind die Studien experimenteller Natur, d.h. Studienteilnehmern werden zufällig in Untersuchungsgruppen aufgeteilt, und zwar in eine oder mehrere Gruppen, die einer oder mehrerer Interventionen unterzogen werden (sogenannte Treatmentgruppen), sowie eine Kontrollgruppe, die zum Zwecke des Vergleichs herangezogen wird und in der Regel keiner Intervention unterzogen wird.

Tabelle 10
Weitere Studieneigenschaften

Schätzmethode	Forschungsdesign		
	Experimentell	Quasi-experimentell	Natürliches Experiment
„Quasi-Differenzen-in-Differenzen“	25	1	0
Varianzanalyse (z. B. ANOVA)	14	4	0
Differenzen-in-Differenzen (DiD)	15	1	0
Kleinste-Quadrate-Schätzungen	10	---	1
Fixed-Effects-Modell	5	1	1
Einfache Querschnittsanalyse	3	---	---
Propensity-Score-Matching	0	1	0
Andere Ansätze	7	1	2
Gesamt	79	9	4

Mehrfachnennungen möglich; die Studien wenden z.T. weitere Methoden an. Die in der Tabelle genannten Methoden beziehen sich auf die in die vorliegende Studie aufgenommenen Schätzer (s. vorangehendes Kapitel)

Neun weitere Studien haben einen quasi-experimentellen Aufbau. Beispiele dafür sind Studien, in denen eine Intervention mit Bewohnern eines Gebäudes stattfindet und hierfür Bewohner eines identischen oder ähnlichen Gebäudes zum Vergleich herangezogen werden. Dadurch, dass die Zuteilung nicht zufällig stattfindet, können strukturelle Unterschiede zwischen den vergliche-

nen Gruppen, in diesem Fall den Bewohnern der unterschiedlichen Gebäude, zu verzerrten Resultaten führen. Durch verschiedene statistische Methoden, z.B. das sogenannte Matching, können diese Verzerrungen jedoch minimiert werden. Drei Labeling-Studien sowie eine Feedback-Studie sind natürliche Experimente, in denen eine „natürliche“, nicht durch den Forscher kontrollierbare Zuteilung in Treatment- und Kontrollgruppe stattgefunden hat, z.B. durch eine unerwartete Politikreform.

Für den Großteil der Studien wird für die Bestimmung der Effekte ein Differenzen-von-Differenzen-Modell (DiD), eine Varianzanalyse oder ein „Quasi-DiD“ genutzt. Bei letzterem handelt es sich um Studien, die zwar kein ökonometrisches Modell mit einem DiD-Schätzer nutzen, die jedoch die gleiche Logik anwenden, nämlich die Differenzen zwischen Treatment- und Kontrollgruppe sowie zwischen Vorher und Nachher heranziehen, um Veränderungen in der Ergebnisvariable zu bemessen (also die Differenzen von Differenzen).

Im Folgenden werden zwei weitere zentrale Studienaspekte genauer betrachtet, die Interventionskombinationen und die Ergebnisvariablen. Dabei wird jeweils zwischen den vier Interventionsstypen und den drei Ergebnisvariablen unterschieden (vgl. Tabelle 5 aus Kapitel 3.4).

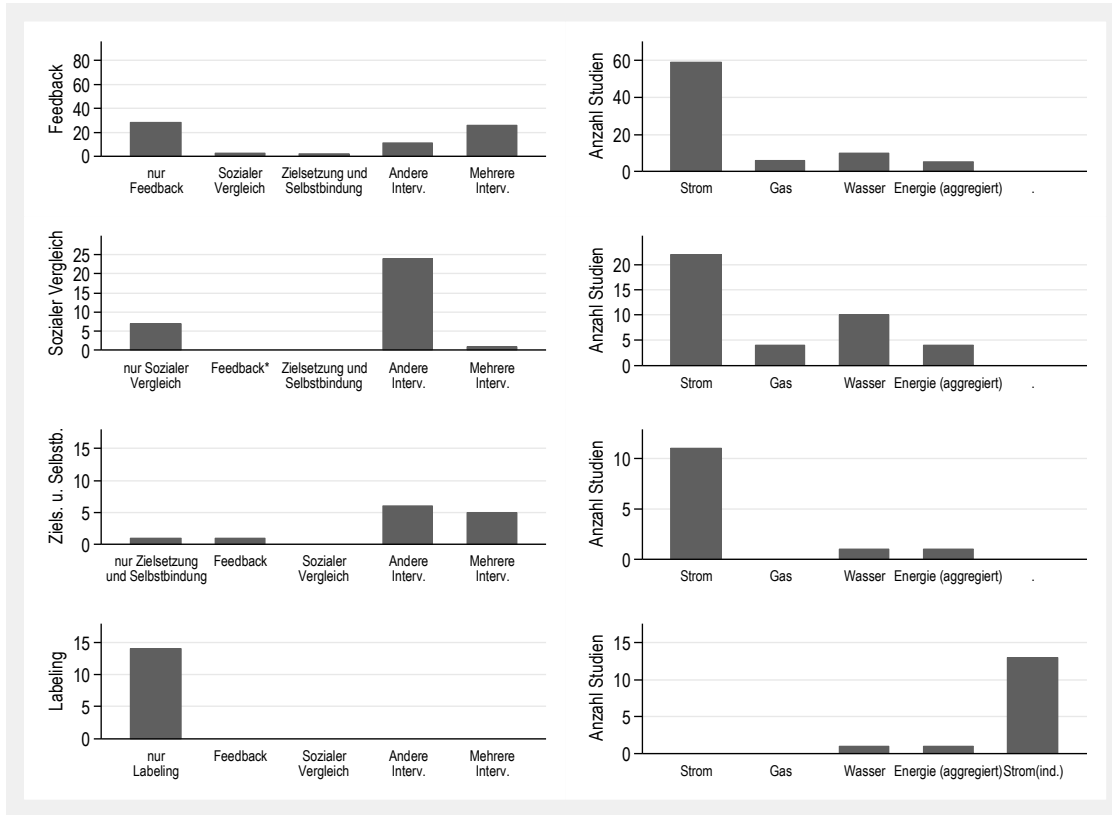
Kombination von Interventionen

In vielen Studien werden Interventionen innerhalb eines Treatments miteinander kombiniert. Für die Isolierung des Effekts einer Intervention stellt dies eine Herausforderung dar. Zudem ist der Vergleich von Effekten unterschiedlicher Studien schwierig, wenn diese nicht dieselben Treatments untersuchen. Abbildung 3 zeigt wie die jeweiligen Interventionen in den Studien miteinander kombiniert werden (linke Spalte). Sofern eine Studie sowohl eine Intervention isoliert als auch in Kombination mit einer oder mehreren anderen anwendet, wird in dieser Grafik angegeben, dass diese Studie die Intervention einzeln betrachtet (z.B. „nur Feedback“). Die Grafik spiegelt also wieder, welche Interventionen bestenfalls isoliert werden können. Unter „Andere Interventionen“ sind insbesondere solche Studien vertreten, in denen die jeweilige Intervention mit Energiespartipps kombiniert wird. Alternativ wurden zusätzlich die Nudges Erinnerungen oder Warnungen ergänzt (vgl. auch Reisch und Sunstein 2017; Sunstein 2014) und in wenigen Fällen wurden die Interventionen mit finanziellen Anreize verknüpft. Labeling wird lediglich in einer Studie mit anderen Interventionen kombiniert, dort jedoch auch als einzelnes Treatment verwendet.

Ergebnisvariablen

Die rechte Spalte von Abbildung 3 zeigt die in den jeweiligen Studien untersuchten Ergebnisvariablen auf. Das Hauptaugenmerk der Primärstudien liegt auf dem Stromkonsum, gefolgt vom Wasser- und Gas-. Einige Studien kombinieren den Strom- und Gasverbrauch, in einem Fall gar mit Kraftstoffverbrauch. Diese sind als „Energie (aggregiert)“ klassifiziert. Bei der Intervention Labeling verhält es sich etwas anders. Da die Berechnung des direkten Effekts von Labeln auf den Energiekonsum schwer ist, bestimmen die meisten Studien indirekte Effekte. Dazu zählen zum Beispiel die Zahlungsbereitschaft für energieeffiziente Geräte und hypothetische Kaufentscheidungen. Entsprechend basieren diese Ergebnisse lediglich auf behaupteten Präferenzen, wohingegen die meisten Studien zu den anderen Interventionen tatsächliche Konsumententscheidungen erfassen. Eine Studie kombiniert Labeling mit Feedback und Sozialem Vergleich. Diese Studie ist die einzige, die den direkten Effekt von Labeln auf den Energieverbrauch untersucht (Kurz et al. 2005).

Abbildung 3
Interventionskombinationen und Ergebnisvariablen nach Intervention



* Die Intervention Sozialer Vergleich beinhaltet stets auch eine Feedbackkomponente. Feedback wird somit als Teil der Intervention und nicht als separate, kombinierte Intervention verstanden.

4.4. Synthese der quantitativen Ergebnisse

Dieser Abschnitt stellt die Ergebnisse der Meta-Analyse entlang der vier Interventionsgruppen Feedback, Sozialer Vergleich, Zielsetzung & Selbstbindung sowie Labeling dar. In den einzelnen Unterkapiteln wird zunächst Strom als am häufigsten anzutreffende Ergebnisvariable präsentiert, ergänzt um Resultate für Gas und Wasser. Die Grundannahme hinter den Interventionen ist zunächst, dass sie sich reduzierend auf den Verbrauch auswirken. Im Sinne einer ergebnisoffenen Analyse wird schlicht jede Abweichung vom „Kontrafaktum“ betrachtet, also von der angemessensten Vergleichssituation, die in experimentellen Studien in der Regel durch die Kontrollgruppe repräsentiert wird.

Für jede Interventions-Ergebnisvariablenkombination wird zunächst eine eigene Meta-Analyse durchgeführt. Die Ergebnisse werden jeweils primär mithilfe sogenannter *Forest Plots* dargestellt. Zum besseren Verständnis dieses klassischen Meta-Analyse-Instruments bietet Anhang A.5 eine kleine Lesehilfe.

4.4.1. Feedback

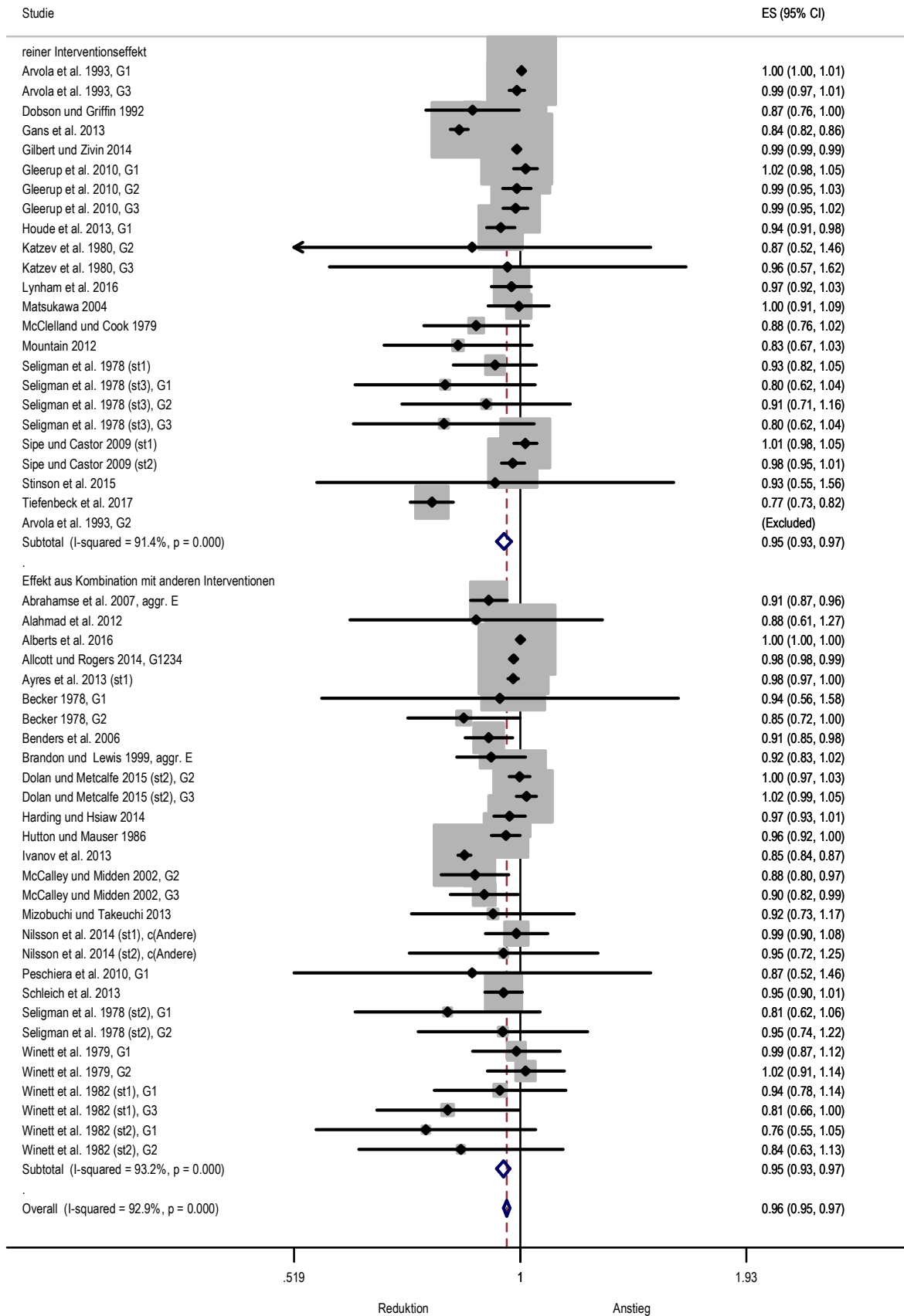
Anhand der Ergebnisse der untersuchten Studien zeichnet sich ab, dass Feedback den Energieverbrauch reduzieren kann. Die Bereitstellung von Informationen über den Verbrauch führt im Schnitt zu einer Reduktion des Stromverbrauchs um vier Prozent ($InRR=0,96$, 95 % KI=0,95-0,97,

$I^2=93\%$, $p=0,00$, 52 Beobachtungen; Abbildung 4). Die Untersuchungen zum Gas- und Wasserverbrauch kommen zu etwas höheren (sechs Prozent) bzw. niedrigeren (drei Prozent) Ergebnissen (Abbildung 5 und Abbildung 6). Ein kleiner Teil der Studien findet positive, jedoch nicht signifikante Ergebnisse. Studien, die Sozialen Vergleich, und damit auch Feedback, untersuchen, werden aus zuovr genannten Gründen nicht in diesem Abschnitt dargestellt, sondern lediglich in Abschnitt 4.4.2. Den stärksten Effekt von Feedback als alleiniger Intervention finden Tiefenbeck et al. (2017). Disaggregiertes Feedback zum Stromverbrauch pro Duschgang, welcher mithilfe einer Anzeige dokumentiert wird, reduziert demnach den Verbrauch um rund 23 Prozent.

Abbildung 4 legt insbesondere nahe, dass es zwei Arten von geschätzten Effekten in der Literatur gibt: zum einen Studien, die relativ unpräzise tendenziell höhere Effekte beobachten (z.B. Katzev et al. 1980, Mountain 2012, Seligman et al. 1978) und zum anderen Studien, die tendenziell geringere Effekte mit einer höheren Präzision ermitteln (die sich in einem stärkeren Gewicht und somit größeren grauen Quadraten in der Grafik bemerkbar machen). Hierzu zählen zum Beispiel Arvola et al. (1993), Lynham et al. (2016), Glerup et al. (2010) und Gilbert und Zivin (2014).

Statistische Tests auf Heterogenität legen starke grundsätzliche Unterschiede zwischen den zugrundeliegenden Studien nahe. Der I^2 -Wert in Abbildung 4 liegt bei 89 bis 93 Prozent, und damit deutlich über dem Wert von 50 Prozent, ab dem von substantieller Heterogenität gesprochen werden kann (Higgins und Green 2011). Es ist also davon auszugehen, dass systematische Unterschiede zwischen den Studien vorliegen, möglicherweise bedingt durch methodische Unterschiede, kontextuelle Unterschiede oder Unterschiede in der konkreten Umsetzung der Interventionen. Die Annahme, dass sich anhand der Literatur tatsächlich ein gepooltes Gesamtergebnis berechnen lässt, ist somit angreifbar. Umso wichtiger ist es, der Heterogenität der Wirkungen auf die Spur zu kommen. Dieser Frage wird das Kapitel 4.5 gewidmet. An dieser Stelle sei jedoch bereits auf ein Spezifikum der Intervention Feedback (und in ähnlicher Weise des im folgenden Kapitel besprochenen Sozialen Vergleichs) hingewiesen: das Medium, über das der Haushalt das Feedback erhält, hat vermutlich einen relevanten Einfluss auf den Interventionseffekt. Während wir diesen Zusammenhang nicht quantitativ untermauern können, stellen wir als Tendenz fest, dass online, via Email oder In-Home-Display (IHD) vermitteltes Feedback einen höheren Effekt als zum Beispiel mittels Briefen zu haben scheint.

Abbildung 4
Effekt von Feedback auf den Stromverbrauch



Eine Erläuterung der abgekürzten Studienattribute findet sich im Anhang A.5.

Abbildung 5
Effekt von Feedback auf den Gasverbrauch

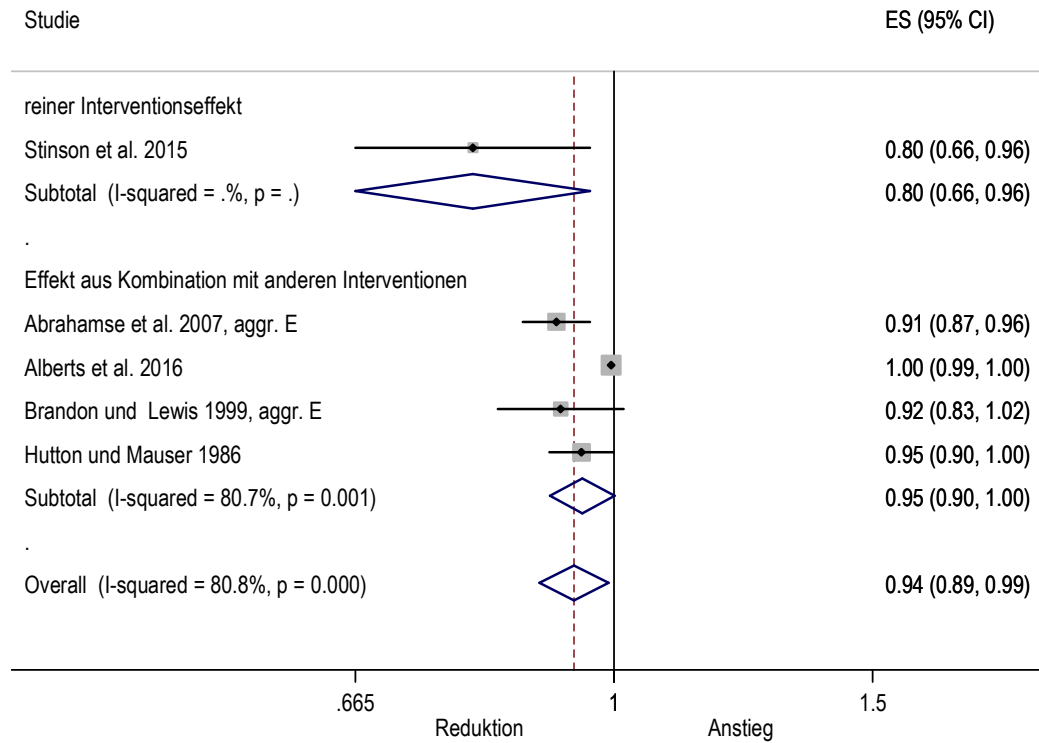
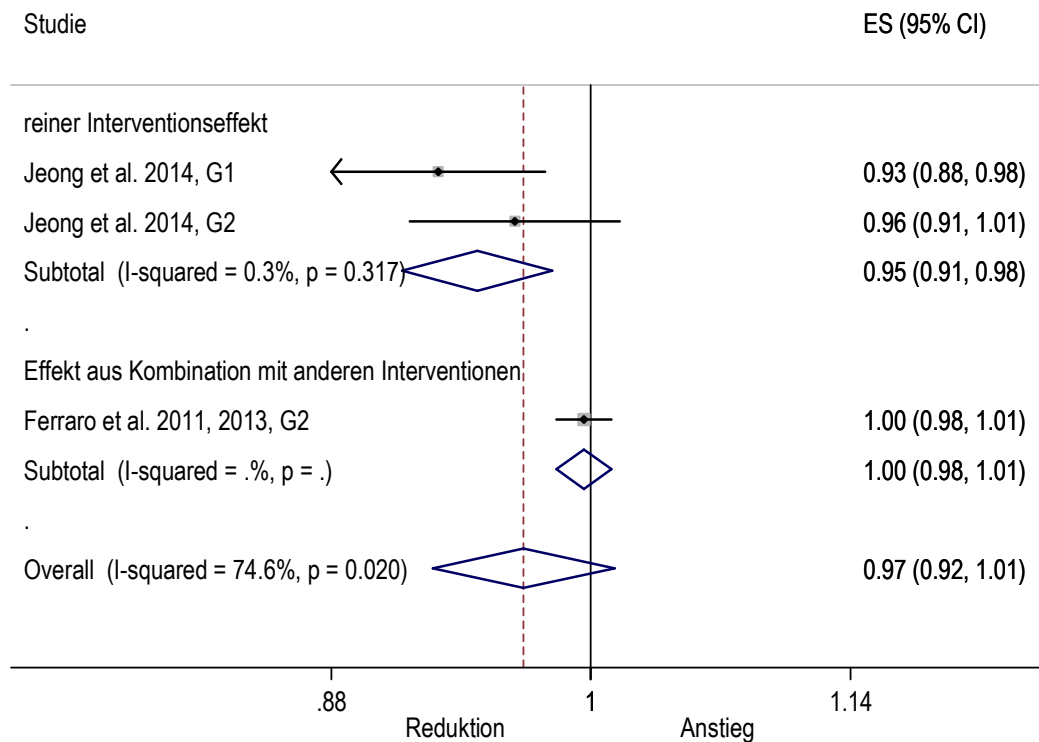


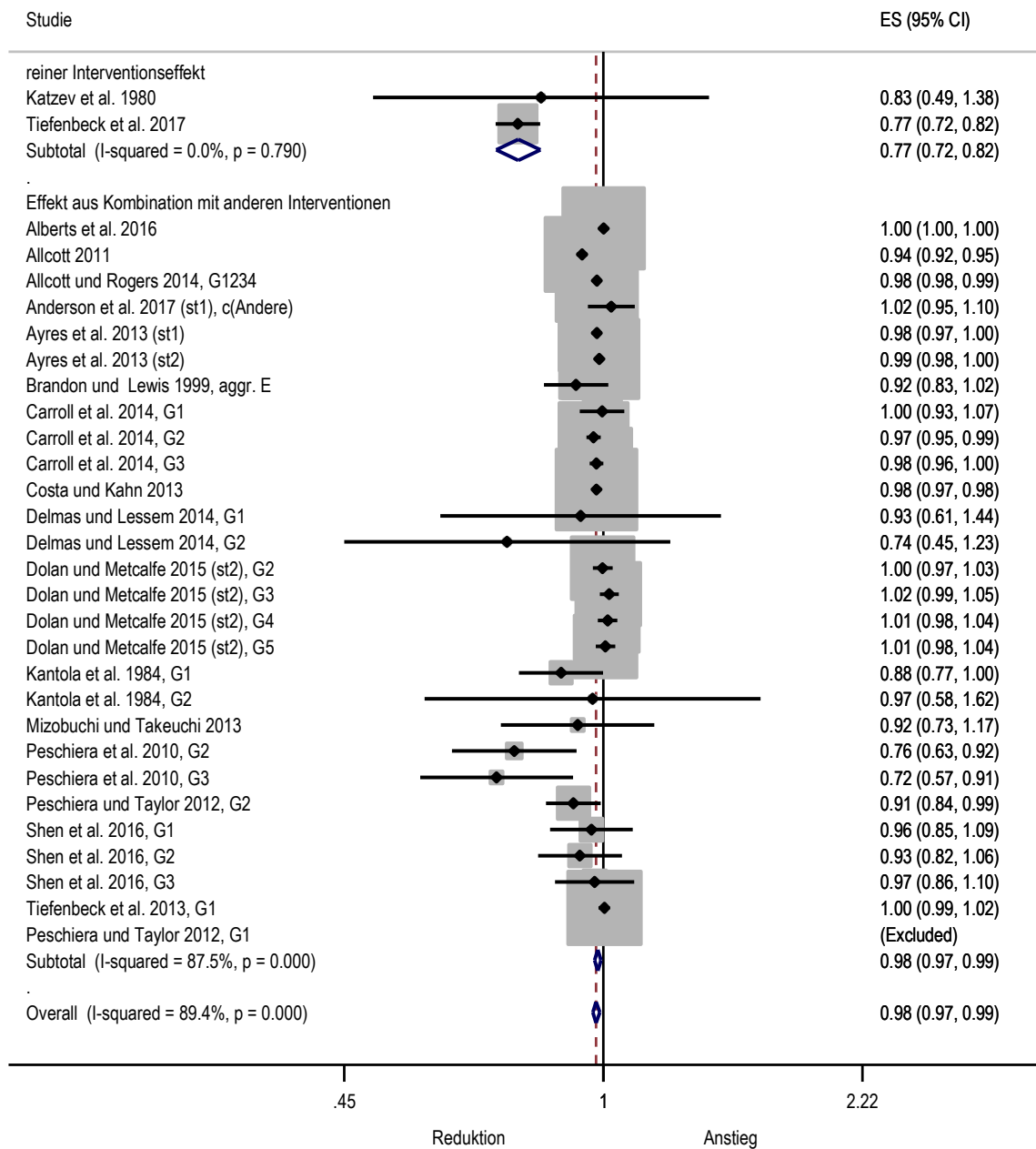
Abbildung 6
Effekt von Feedback auf den Wasserverbrauch



4.4.2. Sozialer Vergleich

Insgesamt deuten die empirischen Studien darauf hin, dass Sozialer Vergleich wirksam darin ist, den Energieverbrauch zu reduzieren. Die vorhandene Literatur legt nahe, dass eine Soziale Vergleichsintervention zu einem mittleren reduzierten Energieverbrauch privater Haushalte von zwei Prozent führt ($InRR=0,98$, 95 % KI=0,97-0,99, $I^2=89\%$, $p=0,00$, 30 Beobachtungen; Abbildung 7). Studien zu Gas- und Wasserverbrauch legen ähnliche Einsparmöglichkeiten nahe, wobei die Anzahl an Studien in diesen Bereichen abermals gering ausfällt (Abbildung 8 und Abbildung 9). Die einzige Kontrollgruppen-Studie, die einen (leichten) Anstieg des Verbrauchs beobachtet, ist Dolan und Metcalfe (2015). Unter den übrigen Studienergebnissen, die eine Verbrauchsreduktion finden, sind Zweidrittel statistisch signifikant.

Abbildung 7
Effekt von Sozialen Vergleichen auf den Stromverbrauch



Eine Erläuterung der abgekürzten Studienattribute findet sich in Anhang A.5.

Abbildung 8
Effekt von Sozialen Vergleichen auf den Gasverbrauch

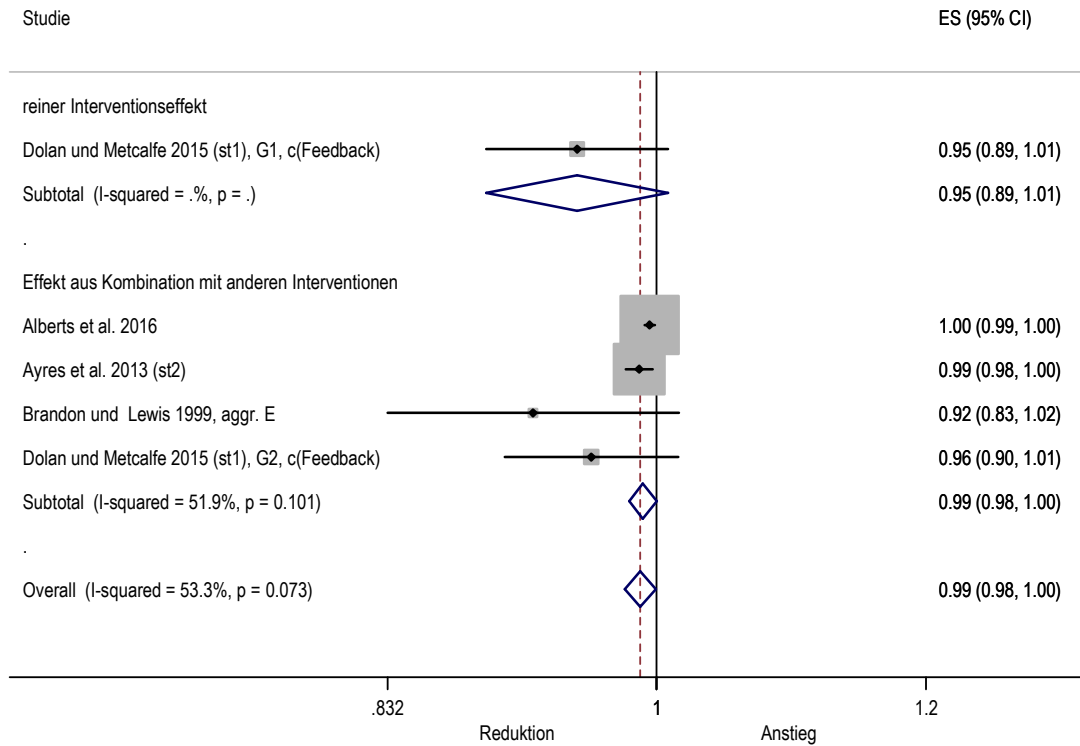
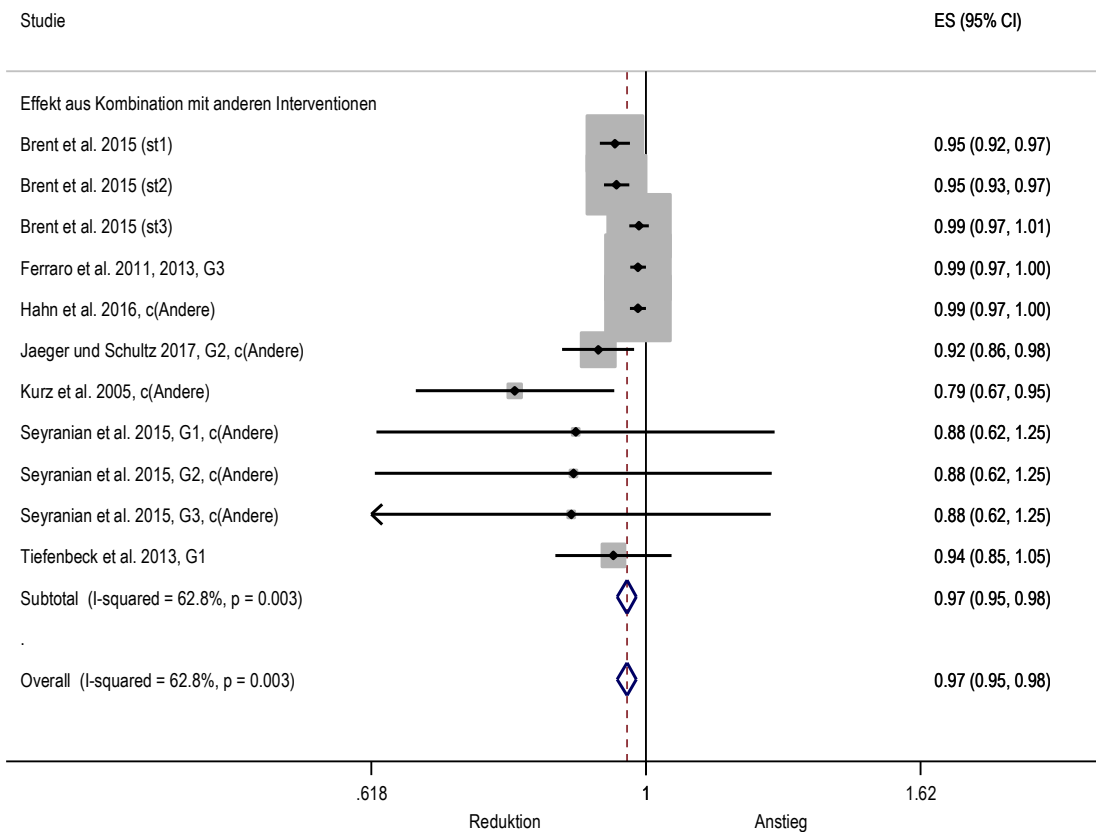


Abbildung 9
Effekt von Sozialen Vergleichen auf den Wasserverbrauch



Eine genauere Betrachtung von Dolan und Metcalfe (2015) zeigt, dass sich der Soziale Vergleich vor allem bei denen als kontraproduktiv erweist, die zum Zeitpunkt der Intervention unter dem Wert ihres Sozialen Vergleichs liegen. Selbiges Phänomen zeigt sich auch bei einer Untergruppenanalyse von Personen mit geringem Konsum in Schultz et al. (2007). Dies kann als sogenannter „Bumerang-Effekt“ interpretiert werden: negative Effekte der Intervention sind bei denjenigen zu beobachten, die vorher besser als die Norm gewesen sind (vgl. Clee und Wicklund 1980). Schultz et al. (2007) zeigen aber auch, dass diesem durch den expliziten Verweis auf injunktive Normen, also auf das von anderen wertgeschätzte Verhalten, begegnet werden kann. Einen ähnlich unerwünschten Nebeneffekt beobachten Tiefenbeck et al. (2013). Sie finden eine signifikante Reduzierung des Wasserverbrauchs um sechs Prozent, bewirkt durch wöchentliches Feedback über den Pro-Kopf-Wasserverbrauch, begleitet von einem Sozialen Vergleich mit den effizientesten 10 Prozent der Nutzer und Tipps zur Einsparung von Wasser. Bemerkenswerterweise stieg gleichzeitig der Stromverbrauch der Treatmentgruppe um 5,6 Prozent. Dieses Ergebnis könnte als erstes Indiz für den sogenannten moralischen Selbstlizenzierungseffekts (*moral licensing*) angesehen werden. Darunter wird das Phänomen verstanden, dass in der Vergangenheit gute Taten eine positive Selbstwahrnehmung begünstigen, die ihrerseits bewirkt, sich eher auf Verhaltensweisen einzulassen, die weniger mit ihren moralischen Werten im Einklang stehen (Nisan und Horenczyk 1990). Treten solche Nebenwirkungen auf, ist es unklar, ob eine Intervention ein nachhaltiges Verhalten im weiteren Sinne induziert, auch wenn der geschätzte Effekt auf eine Reduktion des "direkten" Ergebnisses hindeutet (in diesem Beispiel: Wasserverbrauch).

Als reine Soziale Vergleichsintervention verstehen wir auch Studien, die dem Haushalt Feedback geben, da die Trennung der beiden Interventionstypen eher künstlich erscheint. Dennoch hat eine Studie den Versuch unternommen, den reinen Effekt des Sozialen Vergleichs (ohne Feedback) auf den Energieverbrauch zu identifizieren. Haakana et al. (1997) finden bei reinem Sozialen Vergleich ähnliche oder geringere Effekte als bei einer Kombination der Intervention mit Energiespartipps. Die Studie liefert jedoch keine Informationen über Präzision oder Signifikanzniveaus der berechneten Schätzer, so dass keine standardisierten Werte in die Meta-Analyse aufgenommen werden konnten.⁸

Einige Studien erlauben es, den Effekt zu ermitteln, der damit einhergeht, dass andere Interventionen um eine Komponente eines Sozialen Vergleichs ergänzt werden (Tiefenbeck et al. 2017; Ferraro und Price 2013; Mizobuchi und Takeuchi 2013; Schultz et al. 2015). Die Unterschiede sind größtenteils gering, legen aber nahe, dass sich der Einspareffekt tendenziell durch die zusätzliche Anwendung von Sozialen Vergleichen verstärkt. Schultz et al. (2015) finden gar eine signifikante Reduktion des Energieverbrauchs um neun Prozent im Vergleich zu einem insignifikanten Effekt der ursprünglichen Intervention.

Unter den spezifischen Interventionsdesigns sind die auf Sozialem Vergleich aufbauenden "Home Energy Reports" (HER) des bereits erwähnten privaten Dienstleisters Opower hervorzuheben. Opower kooperiert mit zahlreichen US-Energieversorgern und schickt HER an mehr als zehn Millionen Haushalte in den USA mit dem Ziel, den Stromverbrauch zu senken. Ein HER ist ein zweiseitiger Brief – auf der ersten Seite findet sich ein Balkendiagramm, das den Energieverbrauch des Haushalts mit dem seiner geografisch nächsten Nachbarn in ähnlichen Hausgrößen

⁸ Eine weitere Studie zu Sozialem Vergleich im Energiesparkontext ist Komatsu und Nishio (2015). Die Autoren analysieren zwar kein Verbrauchsverhalten, dafür die Selbsteinschätzung bezüglich des Verbrauchsverhaltens sowie die Motivation zum Energiesparen. Sie finden mehrdeutige Effekte ihrer Sozialen Vergleichsintervention, berichten aber keine Effektgrößen.

vergleicht, sowie Energiespartipps auf der zweiten Seite. Mehrere methodisch hochwertige Studien berichten von signifikanten, jedoch geringen Reduktionseffekten (Allcott 2011b; Ayres et al. 2013; Costa und Kahn 2013; Allcott und Rogers 2014). Mit Ausnahme von Tiefenbeck et al. (2017) sind generell Studien, die größere Effekte finden, mit geringerer Präzision verbunden, wie Abbildung 7 zu entnehmen ist.

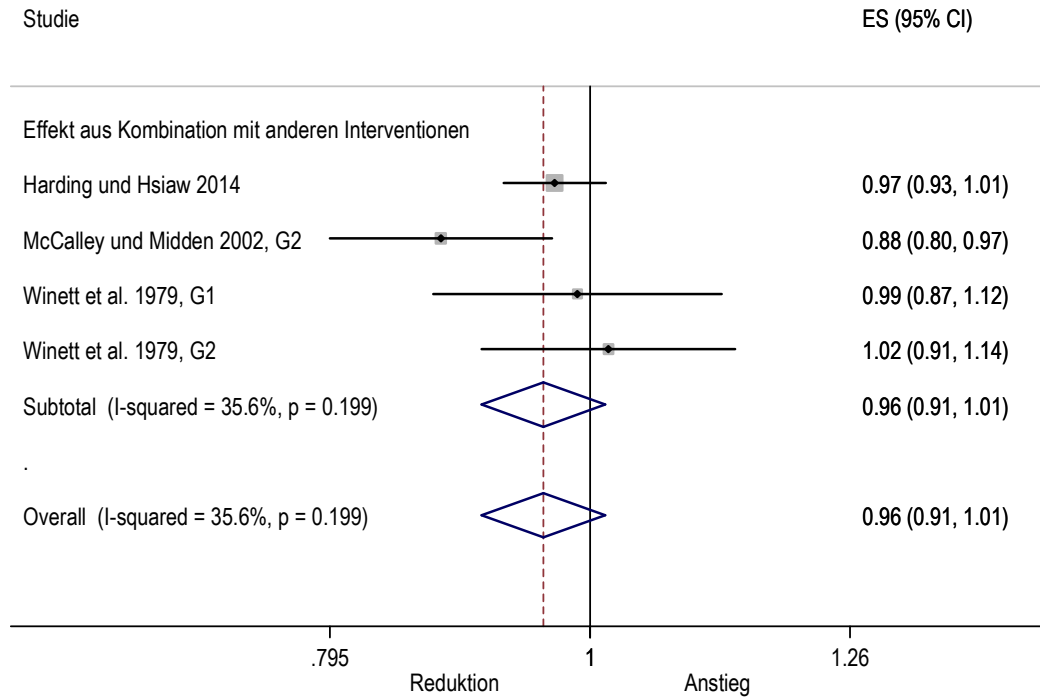
Die konsistent erscheinenden Ergebnisse dürfen jedoch nicht darüber hinwegtäuschen, dass der Test auf Heterogenität mit einem I^2 -Wert von 89 Prozent auch im Falle der Sozialen Vergleichs-Interventionen starke Unterschiede zwischen den zugrundeliegenden Studien nahelegt.

4.4.3. Zielsetzung und Selbstbindung

Während sich die beiden Interventionen Zielsetzung und Selbstbindung als zu ähnlich erwiesen, um getrennt analysiert zu werden, zeigte sich ein anderes relevantes Unterscheidungsmerkmal, anhand dessen wir die Primärliteratur zu diesen Interventionen einteilen: ob das Ziel oder die Verpflichtung selbstgesetzt oder fremdbestimmt ist, da anzunehmen ist, dass ein sich selbst gesetztes Ziel andere Anreize setzt als ein von außen gesetztes.

Zu selbstgesetzten Zielen und Selbstbindung konnten vier Effektgrößen aus drei Studien identifiziert werden (Abbildung 10). Die Ergebnisse sind verhältnismäßig unpräzise geschätzt und weisen somit keinen signifikanten aggregierten Effekt aus ($\ln RR=0,96$, 95 % KI=0,91-1,01, $I^2=36$ %, $p=0,20$, vier Beobachtungen). Harding und Hsiaw (2014) ermitteln anhand einer Feldstudie mit mehr als 12.000 Haushalten, dass selbstgesetzte Ziele insgesamt den Stromkonsum um drei Prozent reduzieren. Unteranalysen dieser Studie kommen zu dem Ergebnis, dass die Gruppe, die ein erreichbares Energieeinsparziel zwischen null und 15 Prozent wählt, ihren Verbrauch sogar um 11 Prozent reduziert. Auch nach 18 Monaten messen die Autoren noch immer einen signifikanten Reduktionseffekt in dieser Untergruppe. Gleichzeitig sparte auch diejenige Gruppe zu Beginn Energie, die ein optimistischeres Ziel (15 bis 50 Prozent) gewählt hatte. Die Energieeinsparanstrengungen in dieser Gruppe verflüchtigten sich jedoch, nachdem sie Feedback über die erzielten Einsparungen erhielten, vermutlich weil ihnen bewusst wurde, dass es unmöglich sein würde, das ehrgeizige Ziel zu erreichen. Im Gegensatz dazu zeigten beide Treatmentgruppen, die entweder ein Null-Ziel oder ein Ziel von mehr als 50 Prozent wählten, keinerlei Verhaltensänderung. Zu diesen Ergebnissen ist jedoch anzumerken, dass sie durch Selbstselektion verzerrt sein könnten: anders als bei experimentellen Studien konnten die Studienteilnehmer selber entscheiden, welche Art des Treatments (Höhe des Einsparziels) sie wünschen, sich also in eine bestimmte Intervention selektieren. Menschen, die bewusster Energie nutzen, entschieden sich beispielsweise vermutlich eher für realistische Ziele.

Abbildung 10
Effekt von selbstbestimmter Zielsetzung und Selbstbindung auf den Stromverbrauch



Nur eine Studie betrachtet den Effekt extern gesetzter Ziele isoliert. Der aggregierte Effekt von sechs Prozent ist nicht signifikant, wobei nicht auszuschließen ist, dass dies auf die geringe Stichprobengröße zurückzuführen ist ($lnRR=0,94$, 95 % KI=0,79-1,11, $I^2=0\%$, $p=0,92$, zwei Beobachtungen, Abbildung 11). Studien, die andere Interventionen mit Zielsetzung kombinieren, sind verbreiteter, sodass eine Effektbestimmung hier verlässlicher ist. Es lässt sich ein übergeordnet signifikanter Effekt beobachten ($lnRR=0,95$, 95 % KI=0,93-0,98, $I^2=43\%$, $p=0,02$, 23 Beobachtungen, Abbildung 11).

In der Studie von Becker (1978) haben sowohl ein geringes als auch ein ambitionierteres Ziel keinen signifikanten Effekt, wohingegen diese in Kombination mit Feedback einen größeren Effekt aufweisen. Auch in der Studie von Seligman et al. (1978) ist die Stromreduktion in der Treatmentgruppe, die ein höheres Einsparziel gesetzt bekommt und Feedback erhält, am höchsten. Winett et al. (1982) haben zwei, zeitlich versetzte RCT-Studien durchgeführt, jedoch mit relativ kleinen Gruppengrößen. In einer der beiden Studien wurden gerade einmal 83 Haushalten auf fünf Treatmentgruppen verteilt. Nach der Teilnahme an einem Treffen motivierten die Forscher die Teilnehmer dazu, sich zu verpflichten, ihren Energieverbrauch innerhalb der nächsten 35 Tage um 15 Prozent zu reduzieren. Neben unterschiedlich häufigem Feedback über ihre Fortschritte erhielten die Haushalte Videobänder mit unterschiedlichen Arten von Tipps zum Energiesparen. Eine Studie wurde im Winter, die andere im Sommer durchgeführt. Alle Treatmentgruppen zeigten signifikante und relativ hohe Reduktionseffekte von sechs bis 24 Prozent.

Unter den Studien zum Gasverbrauch konnten Abrahamse et al. (2007) in ihrem RCT einen neunprozentigen, signifikanten Effekt eines Fünf-Prozent-Einsparziels messen. Zusätzlich zu diesen Erkenntnissen weist eine Laborstudie von McCalley und Midden (2002) auf Energieeinsparpotenziale von extern gesetzten Zielen von rund zehn Prozent hin.

Abbildung 11
Effekt von fremdbestimmter Zielsetzung auf den Stromverbrauch

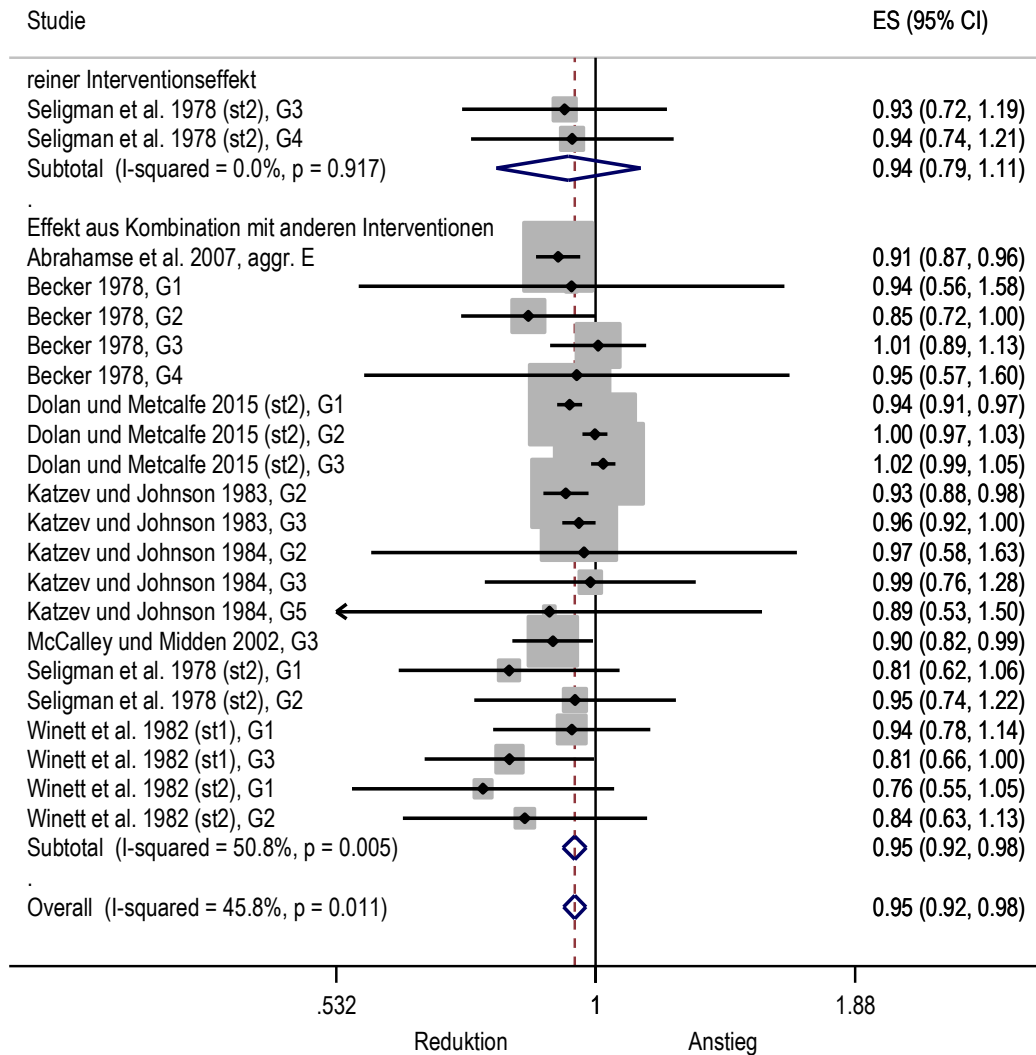


Abbildung 12
Effekt von Zielsetzung und Selbstbindung auf den Gasverbrauch

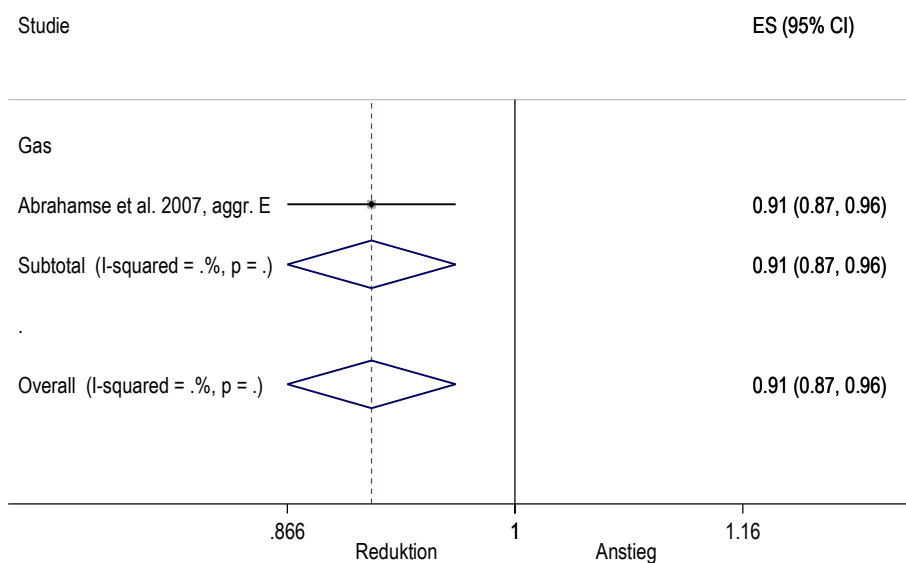
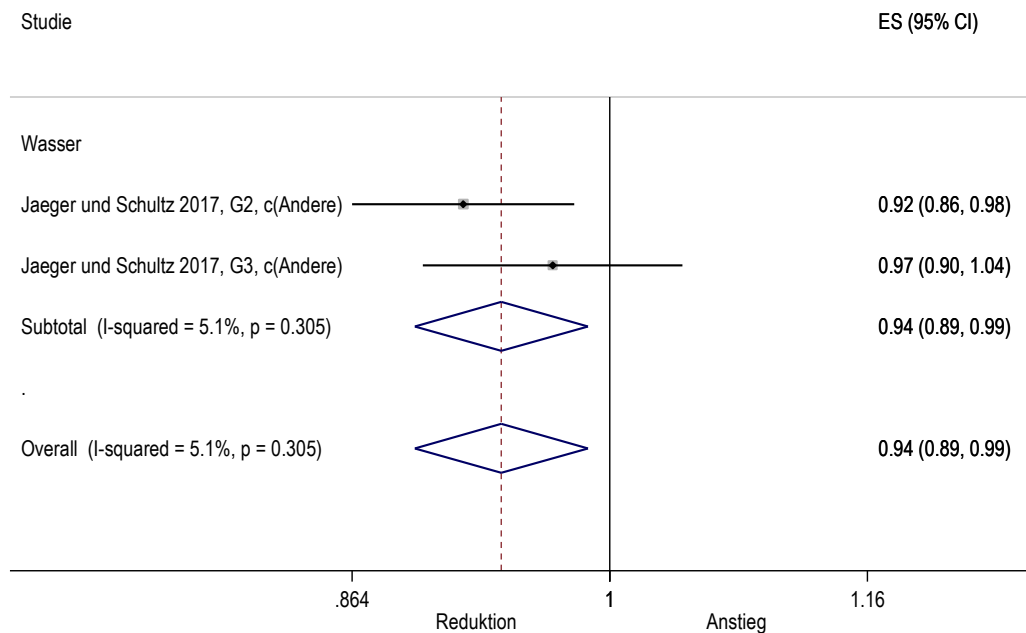


Abbildung 13
Effekt von Zielsetzung und Selbstbindung auf den Wasserverbrauch



Die Werte von Jaeger und Schulz (2017) beziehen sich nur auf diejenigen Haushalte in der Treatmentgruppe, die sich tatsächlich in Form einer Unterschrift selbst an das gesteckte Ziel gebunden haben (somit nur die sogenannten "compliers"). Der Durchschnittseffekt der Intervention in der gesamten Treatmentgruppe ist somit geringer ausgefallen, wurde von den Autoren jedoch nicht berichtet.

4.4.4. Labeling

Für die Intervention Labeling können die Ergebnisse nur für wenige Studien standardisiert werden. Zudem sind die Ergebnisvariablen zu divers, als dass es Sinn machen würde einen übergeordneten Effekt bestimmen zu wollen. Im Folgenden werden die Ergebnisse daher ohne Rückgriff auf meta-analytische Methoden untersucht. Hierbei wird unterschieden zwischen solchen Studien, die Label unter realen Bedingungen analysieren und anderen, die hypothetische Entscheidungen heranziehen, um den Effekt von Labeln zu untersuchen. Die Studien unter realen Bedingungen verfügen vermutlich über eine höhere Übertragbarkeit der Ergebnisse auf andere Kontexte.

Die Mehrheit der identifizierten Labeling-Studien finden signifikante Ergebnisse in mindestens einer Teilstichprobe, hervorgehoben durch die schwarzen Punkte in Abbildung 14 und Abbildung 15. Auch die Effektgrößen dieser Studien deuten auf potenziell starke Effekte von Labeln hin.

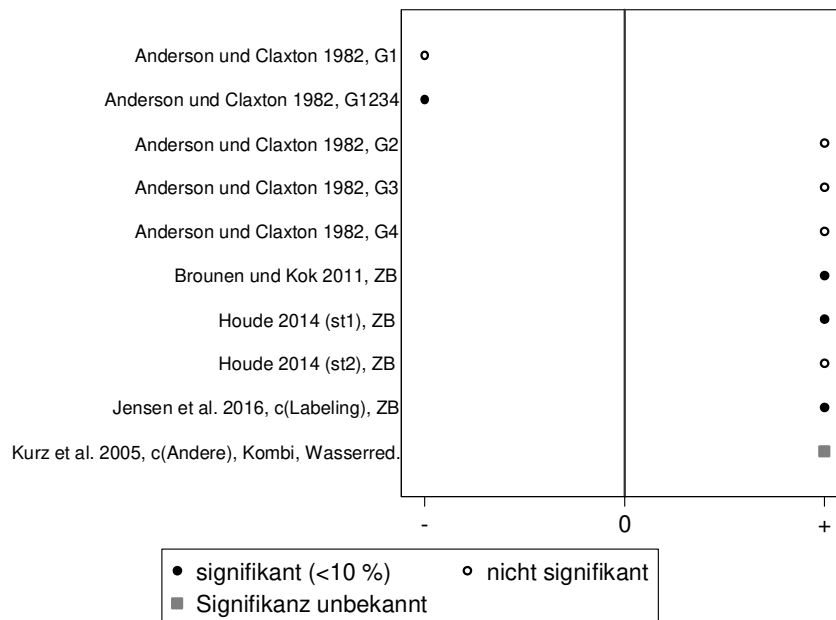
Fünf Studien untersuchen die Wirkung von Labeln unter realen Bedingungen (Abbildung 14). Anderson und Claxton (1982) haben in einem Feldexperiment überprüft wie verschiedene Energielabels den Kühlschrankkauf beeinflussen. Der durchschnittliche Verbrauch gekaufter kleiner Kühlschränke war in Läden mit gelabelten Geräten signifikant niedriger als in den Kontrollläden. Bei großen Kühlschränken konnte kein signifikanter Unterschied gefunden werden. Brounen und Kok (2011) dokumentieren signifikant höhere Verkaufspreise für Häuser in den Niederlanden mit einem "grünen" Energieeffizienzlabel der EU (Klasse A, B, C) als für vergleichbare Häuser ohne ein solches Label. Der Unterschied beträgt 3,6 Prozent. Jensen et al. (2016) zeigen für Dänemark,

dass der Preisaufschlag für Häuser mit guten EU-Energieeffizienzlabels (A bis C) von 2,4 auf 10,1 Prozent gegenüber ansonsten vergleichbaren energieineffizienten Häusern (Labels D bis G) gestiegen ist, nachdem 2010 in Immobilienanzeigen die Angabe von Energieausweisen verbindlich vorgeschrieben wurde. Sowohl Brounen und Kok (2011) als auch Jensen et al. (2016) finden zudem, dass Häuser mit einer schlechten Energieeffizienzklasse (Klasse F und G) im Vergleich zu jenen mit einer durchschnittlichen Klasse (D) zu einem geringeren Preis verkauft werden.

Ähnlich wie Jensen et al. (2016) nutzt Houde (2014) Ausweitungen in Labeling-Regulierungen als natürliche Experimente, um die Effekte des US-amerikanischen *Energy Star* Labels zu untersuchen. Seine Analyse basiert auf einem Vergleich gleicher Kühlschranksmodelle vor und nach der Verschärfung der Kriterien des *Energy Star* Labels. In einer der beiden Studien beobachtet er eine deutlich höhere Zahlungsbereitschaft für Kühlschränke, die dieses Label tragen, in der zweiten Studie ist dieser Effekt nicht signifikant. Aus den Ergebnissen der Studie von Kurz et al. (2005), die der Treatmentgruppe Etiketten über die Verbrauchswerte der haushaltseigenen Geräte zur Verfügung stellte, lässt sich eine Wasserreduktion von rund 21 Prozent bestimmen.

Auch wenn diesen Studien reale Kauf- und Verbrauchsentscheidungen zugrunde liegen, sind die Ergebnisse dieser nicht-experimentellen Studien mit gewisser Vorsicht zu genießen. So wird zum Beispiel bei der Ermittlung der Preisunterschiede zwischen Häusern unterschiedlicher Effizienzklassen für bedeutsame beobachtbare und quantifizierbare Charakteristika (wie Lage, Größe, Baujahr) kontrolliert, um den Effekt des Labels zu isolieren. Dennoch ist schwer auszuschließen, dass der berechnete Effekt nicht doch auch andere Aspekte in der Substanz der Häuser abgreift, was zu einer Über- oder auch Unterschätzung des Labeling-Effektes führen könnte.

Abbildung 14
Quantifizierung der Effekte von Labeling unter realen Studienbedingungen



Die Abkürzung ZB steht für Zahlungsbereitschaft, Wasserred. für Wasserreduktion und Kombi für einen Effekt aus der Kombination von Labeling mit anderen Interventionstypen. Bezüglich der weiteren Studienattribute sei auf die Erläuterungen in Anhang A.5 verwiesen.

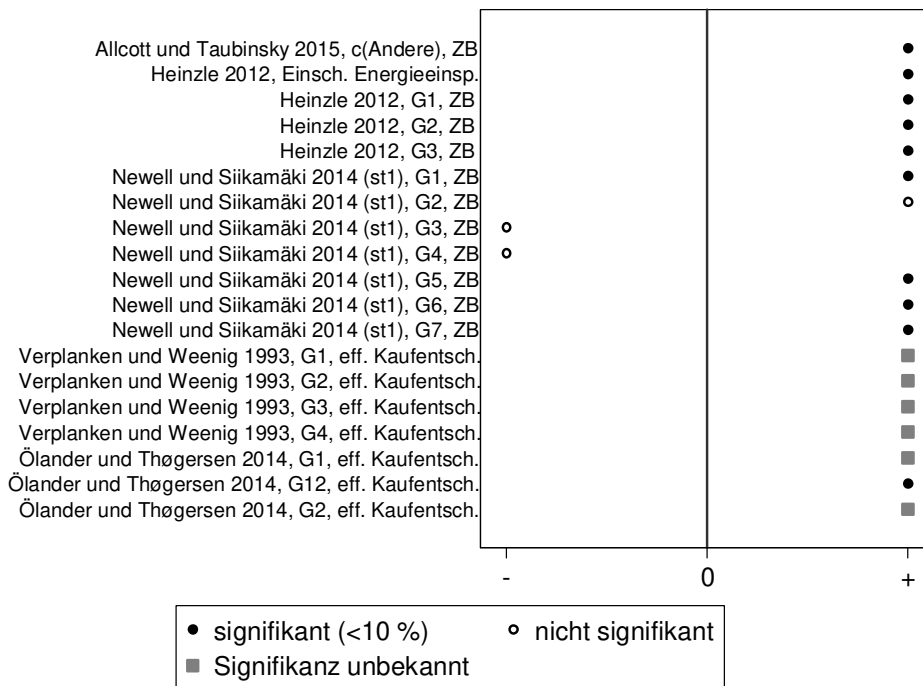
Die Studien, deren Effekte auf hypothetischen Entscheidungen basieren, sind in Abbildung 15 dargestellt. Mehr noch als in der vorherigen Grafik soll angesichts der Unterschiedlichkeit der

angewendeten Maße in Abbildung 15 bloß eine grobe Idee zur Richtung des Effekts vermittelt werden. Die Ergebnisse von Allcott und Taubinsky (2015) deuten auf eine signifikant höhere Zahlungsbereitschaft für Energiesparlampen hin, wenn Informationen über deren Energieverbrauch im Vergleich zu herkömmlichen Lampen bereitgestellt werden. Das Experiment von Heinzle (2012) zeigt, dass die Zahlungsbereitschaft für einen effizienteren Fernseher deutlich zunimmt, wenn ein Label die Verbraucher über die Betriebskosten aufklärt, die über zehn Jahre üblicherweise anfallen. Sie verringert sich jedoch um fast die Hälfte, wenn das Label nur die jährlichen Betriebskosten angibt (in Abbildung 15 sind diese beiden Gruppen als G2 und G3 mit ihrer Zahlungsbereitschaft (in 1000 €) dargestellt). Eine weitere Untersuchung von Heinzle (2012) zeigt außerdem, dass die Energieeinsparpotenziale eines effizienteren Fernsehers besser eingeschätzt werden, wenn die jährlichen Betriebskosten statt des Energieverbrauchs angegeben werden.

Die Studienteilnehmer an einem weiteren hypothetischen Experiment von Newell und Siikamäki (2014) zeigen eine signifikant höhere Zahlungsbereitschaft für geringere laufende Kosten, die durch den Betrieb einer effizienteren Warmwasseraufbereitung erzielt werden, wenn Informationen über Verbrauchskosten oder CO₂-Emissionen bereitgestellt werden (G1, G5-G7 in der Grafik). Die Darstellung eines *EnergyGuides* oder *Energy Star* Labels hat in dieser Studie einen insignifikant negativen Effekt auf die Zahlungsbereitschaft (G3 und G4).

Abbildung 15

Quantifizierung der Effekte von Labeling in hypothetischen Entscheidungssituationen



Die Abkürzung *eff. Kaufentsch.* steht für *effiziente Kaufentscheidung*.

Ölander und Thøgersen (2014) und Verplanken und Weening (1993) belegen darüber hinaus die positive Wirkung der Visualisierung und Vereinfachung von Informationen in Form eines Labels: Die Umstellung einer Energieeffizienz-Skala von einem komplexen "A+++ – D"-System (G2) auf ein einfacheres "A – G" (G1) verdoppelt die Wahrscheinlichkeit, dass ein energieeffizientes Gerät gewählt wird. Unabhängig davon erhört die Klassifizierung eines Fernsehers in die höchste

Energieeffizienzklasse die Wahrscheinlichkeit, dass dieses Gerät gewählt wird, signifikant. Verplanken und Weening (1993) haben untersucht, inwiefern zwei unterschiedliche Darstellungen des Energieverbrauchs die Entscheidung für einen effizienteren Kühlschrank beeinflussen. Kann die Entscheidung ohne Zeitdruck gefällt werden, ist der Anteil derer, die einen effizienten Kühlschrank wählen, bei Bereitstellung eines grafischen Labels (G2) fast doppelt so groß wie bei einer einfachen Information über den täglichen Verbrauch (G1).

4.5. Moderatoranalyse

Die Heterogenitätsanalysen der einzelnen Interventionstypen zeigten auf, dass zwischen den jeweils betrachteten Studien deutliche Unterschiede herrschen, die nicht den üblicherweise beobachtbaren zufälligen Schwankungen zuzuschreiben sind. Moderatoranalysen versuchen dieser Heterogenität auf den Grund zu gehen, indem Variablen untersucht werden, die als vermittelndes Element zwischen den unabhängigen und der abhängigen Variablen wirken können, in unserem Fall den Interventionen und den Energieverbräuchen. So können vor allem methodische und kontextuelle Faktoren eine Rolle spielen, wie ausgeprägt der Effekt einer bestimmten Intervention auf Verbräuche ist. Nicht zuletzt ist die konkrete Implementierung der Intervention von großer Bedeutung. Ein Beispiel eines methodischen Aspektes wurde bereits erwähnt: Studien mit höheren Stichprobengrößen weisen tendenziell kleinere Effekte nach, eine Beobachtung, die auch in früheren Überblicksarbeiten gemacht worden ist (Karlin et al. 2015; Delmas et al. 2013). Im Folgenden sollen weitere mögliche Einflussfaktoren untersucht werden, zunächst anhand einer regressionsbasierten Analyse und dann spezifisch auf die Frage bezogen, ob sich Effektgrößen in der kurzen und langen Frist systematisch unterscheiden.

Metaregression verschiedener potentieller Einflussfaktoren

Tabelle 11 zeigt drei Metaregressionsanalysen, jeweils mit unterschiedlichen Kontrollvariablen, die sämtliche Schätzergebnisse der vorangegangenen Kapitel (mit Ausnahme der Labelingstudien) zusammenfassen. Bei der Zusammenstellung der zugrundeliegenden Stichprobe wurde abermals darauf geachtet, die Unabhängigkeit der aufgenommenen Beobachtungen zu gewährleisten. So wurde in den (wenigen) Fällen, in denen für ein und dieselbe Gruppe unterschiedliche Ergebnisvariablen untersucht wurden, nur das Ergebnis für Strom in die Rechnung mitaufgenommen. Die in den Regressionen untersuchte abhängige Variable ist eine Umformung unserer verwendeten Effektgrößenstatistik $\ln RR$. Diese wurde so umgeformt, dass das Ergebnis die Änderung der Einsparungen widerspiegelt: ein positiver Wert legt nahe, dass der entsprechende Faktor für höhere Einsparungen sorgt, ein negativer Wert reduziert die Einsparungen. Wie in den obigen *Forest Plots* sind die Koeffizienten exponentiert, so dass sie als Prozentveränderung interpretiert werden können. Somit ist beispielsweise ein Koeffizient von 0,05 als fünfprozentige Änderung zu interpretieren.

In der ersten Regression wird zunächst bloß für den Interventionstyp, die Ergebnisvariable und die Kombination mit anderen Interventionen kontrolliert. Die Ergebnisse deuten darauf hin, dass keine oder nur geringe Unterschiede zwischen den untersuchten Interventionen sowie der Ergebnisvariablen existieren.

In die Regression (2) wurden zudem kontextuelle, methodische und implementierungsbezogene Faktoren mitaufgenommen. Die Interventionen scheinen im asiatischen und ozeanischen Raum einen etwas geringeren Einfluss auf den Energiekonsum zu haben, wenngleich dieser Effekt nicht signifikant ist (zweite Spalte). Liegen der Studie automatisch generierte Daten zugrunde, finden wir signifikant um fünf Prozent höhere Einspareffekte, und dass, obwohl diese Daten in der Regel mit einer großen Stichprobenanzahl einhergehen. Eher kontraintuitiv ist zudem, dass Studien, welche die Interventionen mit anderen Interventionen kombinieren, einen

niedrigeren Effekt finden als Studien, die die Hauptinterventionen isoliert betrachten. Bei beiden Aspekten ist jedoch hervorzuheben, dass wir natürlich nur über ein begrenztes Set an Kontrollvariablen verfügen, dass nicht die gesamte Variation in den Daten erklären kann, so dass diese Beobachtungen stärker korrelativen als kausalen Charakter haben.

Tabelle 11

Metaregressionsergebnisse, alle Interventionen und Ergebnisvariablen gepoolt

abhängige Variable:	$\ln(1+(1-RR))$ (Effektgröße)		
	(1)	(2)	(3)
Intervention (Basis: Feedback)			
Sozialer Vergleich	0,00 (0,05)	0,01 (0,50)	0,00 (0,03)
Zielsetzung und Selbstbindung	0,02 (0,89)	0,02 (1,24)	0,02 (0,80)
Zielvariable (Basis: Strom)			
Gas	-0,00 (0,02)	0,01 (0,26)	-0,01 (0,33)
Wasser	-0,00 (0,15)	0,01 (0,46)	-0,01 (0,20)
Intervention kombiniert mit einer der Hauptinterventionen	-0,01 (0,64)	-0,00 (0,15)	-0,014 (0,60)
Intervention kombiniert mit anderer Intervention	-0,03* (1,68)	-0,04** (2,18)	-0,03 (1,32)
Region (Basis: Europa)			
Nordamerika		0,00 (0,24)	
Asien und Ozeanien		-0,03 (0,98)	
Nach 2000 veröffentlicht		-0,01 (0,60)	
Experimentell		0,01 (0,59)	
Daten automatisch generiert		0,05*** (2,81)	
Kontrollgruppe erhielt auch Intervention		0,01 (0,23)	
Veröffentlichung (Basis: Working Paper)			
RePEc Recursive Impact Factor für ökonomische Zeitschriften			-0,01 (0,60)
Konstante	0,06*** (5,23)	0,05* (1,66)	0,07*** (4,20)
Anzahl Beobachtungen	117	117	49
Anzahl der Teilgruppen/ Studien	65/56	65/56	33/29
I ² -Statistik	80 %	77 %	91 %
F-test	0,89	1,23	0,83

* $p < 0,1$; ** $p < 0,05$; *** $p < 0,01$ (statistisches Signifikanzniveau)

Die Regression in Spalte 3 (Tabelle 11) soll den möglichen Einfluss der Qualität des Artikels untersuchen, gemessen an dem *Impact Factor* des Journals, in dem die jeweilige Studie veröffentlicht wurde. Angesehener Journals veröffentlichten in der Regel bessere Studien, die einen höheren *Impact Factor* aufweisen. Da uns kein über die Disziplinen hinweg erhobener *Impact Factor* vorliegt, beschränken wir uns auf ökonomische Artikel als die größte Gruppe in unserem Sample. Für etwa die Hälfte der Studien in dieser Meta-Analyse konnte der verbreitete *RePEc Recursive*

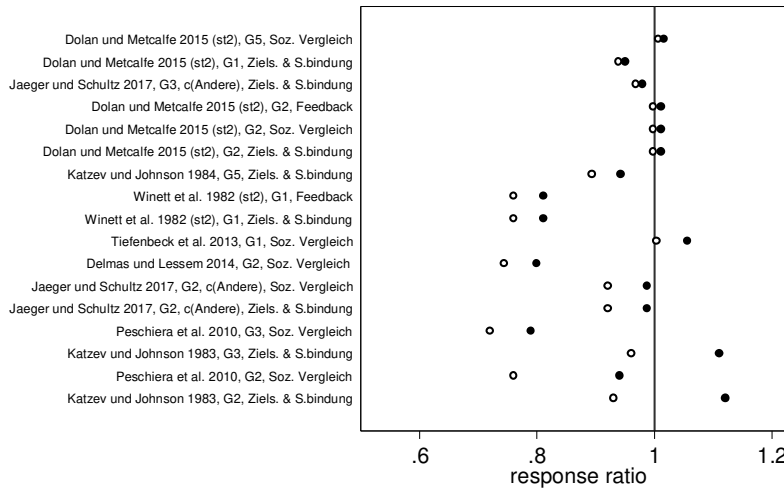
Impact Factor ermittelt werden. Der berechnete Koeffizient ist mit -1 Prozent sowohl statistisch wie ökonomisch insignifikant. Die ermittelten Effektgrößen scheinen somit nicht von der Qualität des Artikels beeinflusst zu sein.

Kurz- und Langfristeffekte

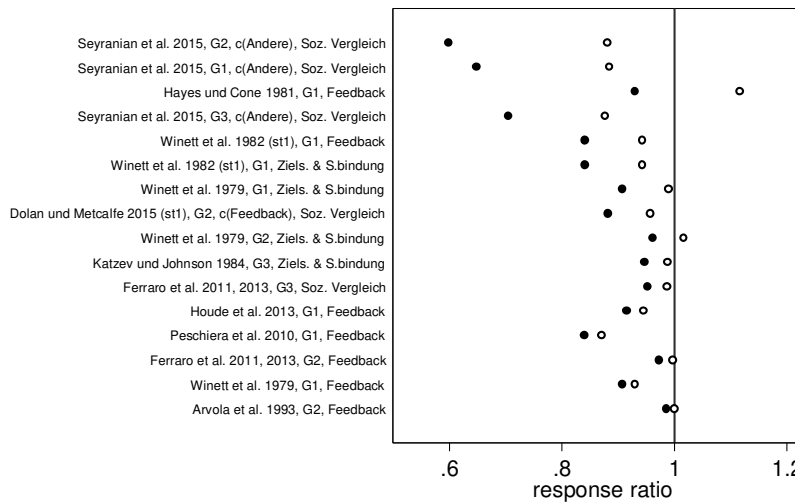
Die Kurz- oder Langfristigkeit der untersuchten Effekte mag zum einen selbst einen moderierenden Einfluss auf Effektgrößen haben. So mag es sein, dass in der kurzen Frist Effekte gefunden werden, die sich im Zeitverlauf abschwächen und somit auf systematische Unterschiede zwischen Kurz- und Langfristuntersuchungen hindeuten. Sie mögen aber auch selbst durch andere Moderatoren beeinflusst sein, z.B. die Dauer der Intervention. Eine Untersuchung der Studien, die sowohl Kurz- als auch Langzeiteffekte ermitteln, ergibt kein klares Bild. Abbildung 16 (a) – (c) stellt die lang- und kurzfristigen *response ratios* der einzelnen Studien gegenüber. Mehrere Studien dokumentieren für einige ihrer Interventionsgruppen eine Verstärkung der Wirkungen auf lange Sicht (Abbildung 16 a). In anderen Studien vermindern sich die Interventionseffekte (siehe Abbildung 16 b), teilweise werden sie fast null (Arvola et al. 1993, Ferraro et al. 2011 & Ferraro und Price 2013). Die in Abbildung 16 (c) dargestellten Studien zeigen in der langen und kurzen Frist dieselben Effekte, mit einem Toleranzbereich von ein Prozent (z. B. Tiefenbeck et al. 2013, Winett et al. 1979, 1982, Delmas und Lessem 2014). Es wäre wünschenswert, die treibenden Faktoren für diese heterogenen Effekte zu identifizieren. Auf der Grundlage der vorliegenden empirischen Erkenntnisse können wir jedoch keine eindeutigen Hinweise feststellen. Erschwert wird die Analyse dadurch, dass es nachvollziehbarerweise Unterschiede in der konkreten Definition von kurzer und längerer Frist zwischen den Studien gibt, die nicht ausreichend dokumentiert sind. Da Langzeiteffekte für die Kostenwirksamkeit der Interventionen von entscheidender Bedeutung sind, sollte ihre Analyse ein Schwerpunkt der zukünftigen Forschung sein.

Abbildung 16
Vergleich der Effekte in der kurzen und längeren Frist

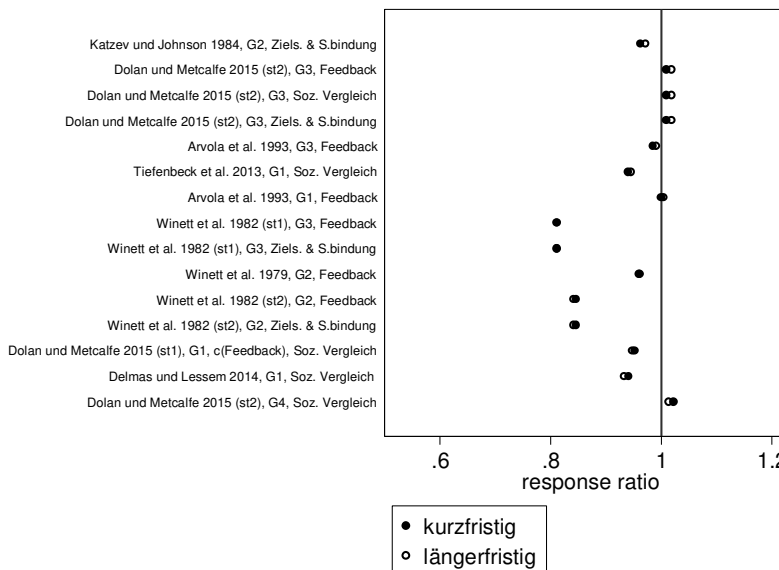
(a) HÖHERE EINSPARUNG IN LANGER FRIST



(b) GERINGERE EINSPARUNG IN LANGER FRIST



(c) KEIN UNTERSCHIED ZWISCHEN EINSPAREFFEKTEN IN KURZER UND LANGER FRIST



5. Zusammenfassung, Schlussfolgerungen und Handlungsempfehlungen

5.1. Zusammenfassung

Die Europäische Union, wie auch viele einzelne Nationen, verfolgen das Ziel, den Energieverbrauch drastisch zu reduzieren. Es stellt sich dabei die grundsätzliche Frage, wie diese starken Energieverbrauchsreduktionen erzielt werden können. Bei der Beantwortung dieser Frage rücken zunehmend Haushalte als Verbraucher sowie die Verhaltensökonomik als Inspiration für das Design politischer Maßnahmen in den Fokus (Ernst et al. 2017, Reisch 2013). Diese Entwicklung basiert auf der Hoffnung, dass auf Haushalte abzielende verhaltensökonomische Interventionen kosteneffiziente Alternativen zu den klassischen Maßnahmen wie Steuern oder Verboten darstellen (Allcott und Mullainathan 2010).

Die vorliegende Studie untersuchte vor diesem Hintergrund, welche Potenziale verhaltensökonomische Ansätze bei der Reduzierung des Energieverbrauchs von privaten Haushalten aufweisen. Betrachtet werden hierbei vier verhaltensökonomische Interventionen, die sich hinsichtlich ihrer Umsetzbarkeit als am besten geeignet erwiesen und in der Literatur am häufigsten untersucht wurden: Feedback, Sozialer Vergleich, Zielsetzung und Selbstbindung sowie Labeling. Für jede dieser vier Interventionstypen schätzen wir auf Basis der Gesamtheit der bisher vorliegenden wissenschaftlichen Studien ab, wie viel Energie durch diese eingespart werden kann. Ein wichtiges Alleinstellungsmerkmal der Studie ist, dass wir dabei nur Primärstudien einbeziehen, die aufgrund ihrer angewandten Methode *kausale* Effekte identifizieren können.

Die bestehende empirische Literatur wurde in Form einer Meta-Analyse systematisch analysiert. Nach einer strukturierten Suche wurden insgesamt 87 Studien einbezogen. Der Hauptteil dieser Studien untersucht die Wirkung auf den Stromverbrauch. Im Durchschnitt führen Feedback, Sozialer Vergleich sowie Zielsetzung und Selbstbindung zu einer Reduzierung des Stromverbrauchs von zwei bis vier Prozent. Diese Effektspannbreite ist für alle vier Interventionstypen zu beobachten und wird prinzipiell durch die Studien zum Gas- und Wasserverbrauch bestätigt, doch liegt hierzu deutlich weniger Evidenz vor.

Eine Besonderheit dieser Studie besteht außerdem darin, dass sie die erste ist, welche die Intervention Labeling in einer Metaanalyse zusammenfasst. Die existierenden Studien erlauben für Labeling zwar keine Berechnung des zu erwartenden Durchschnittseffektes. Zusammenfassend lässt sich aber sagen, dass Labeling durchgehend Potenzial aufweist: Nicht nur hypothetische Kaufentscheidungen in Auswahlversuchen bestätigen die Wirksamkeit von Labels, sondern auch Evaluierungen im Feld.

Die Ermittlung von Durchschnittseffekten wird vielfach dadurch erschwert, dass die jeweilige Intervention nur in seltenen Fällen isoliert betrachtet, sondern stattdessen häufig kombiniert werden. Heterogenitätsanalysen weisen darauf hin, dass größere Unterschiede zwischen Ergebnissen von Studien für den gleichen Interventionstypen bestehen. Dies legt nahe, dass der jeweilige Studienkontext, aber auch die konkrete Ausgestaltung einer Intervention, starken Einfluss auf die Ergebnisse hat. Jedoch lassen sich auf Basis der existierenden Literatur keine „Makromuster“ erkennen, zum Beispiel in Bezug auf regionale Unterschiede, dem Publikationsjahr oder der Qualität des Journals. Wir können somit zum Beispiel keine Hinweise geben, dass etwa die in Europa gemessenen Effekte tendenziell geringer sind als die in den USA. Dennoch scheinen methodische Aspekte die Ergebnismuster zu beeinflussen: Wie auch Delmas et al. (2013) finden wir, dass höhere Einspareffekte tendenziell unpräziser gemessen werden, beispielsweise aufgrund einer kleinen Stichprobe. Dies bedeutet, dass gerade die Studien, die nur kleine Effekte messen, dies präziser und damit „vertrauenswürdiger“ machen.

5.2. Schlussfolgerungen

Die Landesregierung von Baden-Württemberg hat sich im Rahmen der Energiewende das ehrgeizige Ziel gesetzt, bis 2050 seinen Energieverbrauch im Vergleich zu 2010 zu halbieren. Um dieses ambitionierte Ziel zu erreichen, stehen der Verbraucherpolitik neben den klassischen Instrumenten eine Reihe von Maßnahmen, die auf Erkenntnissen aus der Verhaltensforschung basieren, zur Verfügung (Reisch und Sunstein 2017, S. 354).

Die Ergebnisse der Meta-Analyse zu den einzelnen Maßnahmen legen zunächst den Schluss nahe, dass alle untersuchten verhaltensökonomischen Interventionen den Energieverbrauch im Durchschnitt reduzieren, allerdings sind die zu erwartenden Effekte mit durchschnittlich bis zu vier Prozent relativ gering. Dieses Einsparungspotenzial ist außerdem womöglich als (zu) optimistisch einzuschätzen, da nicht auszuschließen ist, dass die Literatur einem sogenannten *Publication Bias* unterliegt.⁹ Hiermit ist die Tendenz gemeint, dass diejenigen Studien eher veröffentlicht werden, die höhere Einsparungen finden, während Studien mit geringeren, insignifikanten Einsparungen (oder gar Verbrauchsanstiegen) seltener publiziert werden (Sterne et al. 2011).

Um eine abschließende Bewertung zu den Interventionen abzugeben, bedarf es – neben der Analyse der Effektivität der Maßnahmen – zusätzlich mindestens einer Kosten-Nutzen-Betrachtung, letztlich sogar einer umfassenden Wohlfahrtsanalyse. Nur wenige der betrachteten Studien liefern spezifische Informationen zu Kosten der Interventionen, die für eine Analyse der Kosteneffektivität genutzt werden können.¹⁰ Auch darüber hinaus gibt es kaum Studien, die die Kosteneffektivität der Maßnahmen betrachten, weshalb eine eindeutige Einschätzung zu diesem Punkt schwierig ist.

Grundsätzlich lassen sich jedoch zwei Sichtweisen in Bezug auf eine Analyse der Kosteneffektivität unterscheiden: die absolute und die relative Kosteneffektivität. In beiden Fällen wird ein Vergleich der Kosten für die Bereitstellung der Maßnahme mit dem erzielten Nutzen angestellt, wobei sich der Nutzen vor allem in der Reduktion der volkswirtschaftlichen Kosten von Treibhausgasemissionen, insbesondere Kohlendioxid (CO₂), bemisst (Nordhaus 2014, IAWG 2013). Bei der *absoluten Kosteneffektivität* wird untersucht, ob dieses Verhältnis für die verhaltensökonomischen Maßnahmen positiv ist, sie also mehr Nutzen stiften als sie kosten. Allcott und Mullainathan (2010) und Allcott (2011b) kommen beispielsweise zu dem Schluss, dass relativ günstige Energiesparbriefe mit einem Sozialen Vergleich (Kosten von etwa 4\$ pro Jahr) im Durchschnitt den Energieverbrauch von US-Haushalten um etwa zwei Prozent senken und dort somit kosteneffektiv sind. Andor et al. (2017) zeigen jedoch die Kontextabhängigkeit dieses Ergebnisses auf. Zum einen ermitteln sie, dass die Maßnahme für Deutschland nicht kosteneffektiv ist. Zum anderen legen sie dar, dass diese Intervention und Maßnahmen mit ähnlichen Einspareffekten – unter anderem aufgrund geringerer Energieverbräuche von privaten Haushalten (etwa 12.300 kWh in den USA vs. 3.300 kWh in Deutschland) – vermutlich in vielen Ländern nicht kosteneffektiv sein können. Basierend auf dieser Informationen und der in dieser Meta-Analyse ermittelten Spannweite von bis zu vier Prozent Einspareffekten, schlussfolgern wir, dass die Maßnahmen in

⁹ Eine asymmetrische Verteilung von Effektschätzern und den dazugehörigen Standardfehlern wird in der Literatur als ein Indiz für *Publication Bias* gewertet (Sterne et al. 2011). Entsprechende Analysen der zugrundeliegenden Studien deuten auf eine solche Asymmetrie hin.

¹⁰ Allcott (2011b), Ayres et al. (2013), Bekker et al. (2010) und Brent et al. (2015) geben beispielsweise den Geldbetrag pro eingesparter Einheit (kWh bzw. Gallone) an. Einige andere Studien stellen die jeweilige Intervention lediglich als günstig oder potenziell kosteneffektiv dar ohne dies quantitativ zu untermauern.

Deutschland voraussichtlich nicht kosteneffektiv sein werden. Diese Einschätzung kann sich jedoch insbesondere durch eine fortschreitende Digitalisierung ändern, welche sowohl die Kosten als auch die Einspareffekte substantiell verändern kann.

Falls die Politik ihr Handeln hingegen strikt daran orientiert, ein gestecktes Energieeinsparziel zu erfüllen, ist die *relative Kosteneffektivität* entscheidend. Hierbei wird die absolute Kosteneffektivität der verhaltensökonomischen Maßnahmen der absoluten Kosteneffektivität von möglichen alternativen Energiesparmaßnahmen gegenübergestellt. Ist somit Energiesparen „an sich“ das Ziel und nicht Mittel zum Zweck der Vermeidung von Treibhausgasemissionen, werden solche Maßnahmen gesucht, die eine Einheit Energie (bspw. beim Strom kWh) am kostengünstigsten vermeiden können. Hier können trotz der kleinen Effekte die Maßnahmen *relativ* kostengünstig sein. Um dies abzuschätzen, bedarf es integrierter Evaluationen von verhaltensökonomischen und alternativen Maßnahmen. Beispielsweise vergleichen Benartzi et al. (2017) die Effekte von klassischen Instrumenten mit denen von verhaltensökonomischen Maßnahmen für eine Reihe von Anwendungsfeldern und kommen zu dem Schluss, dass letztere häufig relativ kostengünstig sind. Sie weisen aber auch darauf hin, dass es mehr Studien mit Kosten-Nutzen-Analysen bedarf, um ein abschließendes Urteil fällen zu können. Basierend auf unseren Ergebnissen muss zudem angemerkt werden, dass die relativen Vergleiche kontextabhängig sind. So zeigt unsere Analyse, wie heterogen die Ergebnisse für denselben Interventionstypen sind, was sich zwangsläufig auch auf die relativen Vergleiche überträgt. Als Schlussfolgerung daraus, müssen integrierte Evaluationen im konkreten Kontext (d.h. Zielregion, Zielpopulation) durchgeführt werden.

Für die Verbraucherpolitik in Baden-Württemberg ziehen wir basierend auf diesen allgemeinen Schlussfolgerungen die folgenden Schlüsse: Aus dem Instrumentenkasten der Verbraucherpolitik erscheinen die vier verhaltensökonomischen Maßnahmen Feedback, sozialer Vergleich, Zielsetzung und Selbstbindung sowie Labeling grundsätzlich als vielversprechende Kandidaten zur Reduzierung des Energieverbrauchs von privaten Haushalten. Die Meta-Analyse offenbart, dass alle vier Maßnahmen durchschnittlich zu Einspareffekten führen, die jedoch mit unter fünf Prozent relativ klein sind. Bei solch geringen Einsparungen erweisen sich die Maßnahmen eher nicht als kostengünstiges Instrument zum Klimaschutz. Weil sich jedoch die Politik ein konkretes Energieparziel gesetzt hat, ist der relative Vergleich der Kosteneffektivität mit alternativen Maßnahmen entscheidend. Da die Meta-Analyse zudem gezeigt hat, dass die Effekte der einzelnen Studien sehr heterogen sind sowie weitere Literatur die Kontextabhängigkeit der Kosteneffektivität aufzeigt und für Baden-Württemberg keine Studienergebnisse zu diesen Interventionen vorliegen, mögen spezifische Maßnahmen sich im Einzelfall als kosteneffektiv erweisen.

5.3. Handlungsempfehlungen

Für die Verbraucherpolitik in Baden-Württemberg bedeutet dies, dass die Maßnahmen systematisch in Baden-Württemberg getestet werden sollten. Alternative Maßnahmen, gegenüber welchen die verhaltensökonomischen Interventionen abgewogen werden müssten, sind insbesondere Preispolitiken (wie beispielsweise eine Erhöhung der Energiesteuern oder auch die Anwendung von dynamischen Preisen) und Standards sowie Verbote.¹¹

¹¹ Eine noch weitergehende innovative Möglichkeit zur Reduktion des Energieverbrauchs stellt die Sektor-*kopplung* dar (acatech et al. 2017). So könnte der Wärme- und Stromsektor gekoppelt werden, um die Stromerzeugung aus erneuerbaren Energien auch für den Wärmesektor zu nutzen. Eine wichtige Voraussetzung hierfür scheint die Revision des Strompreises (bzw. seiner gesetzlich vorgegebenen Bestandteile)

Die Autoren dieser Studie plädieren für eine systematische Evaluation von möglichen Energieeinsparmaßnahmen in Baden-Württemberg durch eine Zusammenarbeit von Wissenschaft, Industrie, Politik sowie relevanten Akteuren wie den Verbraucherzentralen.¹² Dass eine solche Zusammenarbeit nicht nur Wunschdenken ist, sondern auch praktisch umsetzbar und einen Mehrwert für alle beteiligten Seiten hat, zeigen erste Beispiele. Hier sei beispielhaft auf das Projekt „Energiesparen in privaten Haushalten – ein Randomized Controlled Trial (RCT) zur Wirkungsevaluierung einer Endverbrauchersensibilisierung“ hingewiesen, bei dem in Zusammenarbeit zwischen RWI und Energieversorgern wie E.ON Energie Deutschland GmbH, Städtische Werke AG und WEMAG AG sowie der EnergieAgentur.NRW und der Verbraucherzentrale NRW e.V. die Wirkung von verschiedenen Energiesparbriefen auf den Stromverbrauch von Haushalten untersucht wurde (RWI 2017).

Solche systematischen Evaluationen sind für alle potenziellen Maßnahmen zur Reduktion des Energieverbrauchs vorstellbar. Wie sich auch im Rahmen dieser Studie gezeigt hat, ist es hierbei wesentlich, dass die Evaluationen durch ihre Konzeption tatsächlich *kausale* Effekte der entsprechenden Maßnahme identifizieren, die Ergebnisse transparent und umfassend berichten und es ermöglichen, Langzeiteffekte nachzuzeichnen. Insbesondere bietet sich die Evaluation mittels sogenannter kontrollierter, randomisierter Studien (RCTs) an. Hierzu sollten die Maßnahmen in Pilotprojekten unter Zusammenarbeit der Partner (insbesondere Wissenschaft und Industrie) erprobt werden. Gerade die Einführung von Smart Metern, welche in der nahen Zukunft vermehrt vonstattengehen wird, bietet hervorragende Möglichkeiten, um sowohl Preispolitiken als auch verhaltensökonomische Interventionen im Vergleich zu testen. Erst nach der Evaluation und basierend auf der robusten Evidenz sollte die Politik dann entscheiden, welche Maßnahmen großflächig eingeführt werden.

Im Rahmen dieser Evaluationen sollte nicht nur eine reine „klassische“ Kosten-Nutzen-Betrachtung durchgeführt werden, sondern explizit durch ein entsprechendes Forschungsdesign eine Gesamtwohlfahrtsanalyse ermöglicht werden. Dabei gilt es, neben den direkten monetären Kosten und Nutzen auch indirekte Kosten und Nutzen zu berücksichtigen. Gerade bei verhaltensökonomischen Interventionen werden versteckte Kosten (*hidden cost*) vermutet und in ersten Studien auch nachgewiesen (Damgaard und Gravert 2018). So kann ein Sozialer Vergleich zum Beispiel zu psychologischen Kosten führen. Tatsächlich weisen Allcott und Kessler (2017) nach, dass ein Teil der Haushalte bereit ist, dafür zu zahlen, dass sie in der Zukunft keine Energiesparbriefe mehr erhalten. Gleichzeitig ist auch ein zusätzlicher psychologischer Nutzen der Interventionen denkbar, der in den Energie- und Kosteneinsparungen nicht wiedergespiegelt wird. Eine systematische Evaluation sollte daher den Anspruch verfolgen, durch ein intelligentes Forschungsdesign eine umfassende Wohlfahrtsanalyse zu erlauben.

Die Rolle der Politik läge darin, Anreize zu setzen, damit derartige systematische Evaluationen von Maßnahmen und die Zusammenarbeit von Wissenschaft und Industrie stattfinden. Industrieunternehmen (wie Energieversorger oder Einzelhändler) haben zunächst naheliegender Weise wenig Interesse an Projekten, die nicht ihr Kerngeschäft betreffen und sogar zu einer Verminderung ihrer Erlöse führen können. Die Autoren dieser Studie schlagen daher vor, solche gemeinsamen Evaluationsprojekte durch ein finanziell gefördertes Forschungsprogramm oder in

zu sein (Ernst et al. 2017). Auch dies sollte im Rahmen von einer Pilotstudie vor einer Umsetzung systematisch evaluiert werden.

¹² Dieser Aufruf folgt dabei generellen Apellen für evidenzbasierte Politikberatung durch zahlreiche Wissenschaftler (siehe bspw. Boockmann et al. 2014 und Schmidt 2014).

Form einer öffentlichkeitswirksamen Unterstützung anzuregen, je nach Möglichkeiten und Prioritäten der Politik.

Für alle Maßnahmen gilt, dass die Effekte der Maßnahmen in Laborexperimenten und Reallaboren getestet werden können. Allerdings birgt diese Art der Forschung Ungewissheiten bezüglich der externen Validität der Ergebnisse, sprich, ob sich die Effekte auch bei einer großflächigen Anwendung in Baden-Württemberg tatsächlich einstellen würden. Gründe hierfür sind unter anderem, dass sich die Teilnehmer üblicherweise selbst in die Studien selektieren, was zu Problemen bei der Repräsentativität führt, sowie das künstliche Umfeld im Labor. Zudem stellt sich bei Reallaboren, die ohne Kontrollgruppe arbeiten, die Frage, wie der kausale Einspareffekt valide ermittelt werden soll. Obwohl solche Forschung interessante Erkenntnisse liefern kann, plädieren wir für die Einbeziehung rigoroserer Ansätze bei der Evaluation der Maßnahmen.

Konkret bedeutet dies folgendes für die Verbraucherpolitik in Baden-Württemberg: Für die Maßnahmen Feedback, Sozialer Vergleich sowie Zielsetzung und Selbstbindung sollte die Verbraucherpolitik Baden-Württemberg Anreize schaffen, dass Wissenschaftler mit Energieversorgern in gemeinsamen Feldexperimenten die Energieeinspareffekte der Maßnahmen identifizieren. Der wohl bedeutendste Vorteil einer solchen Zusammenarbeit ist, dass das reale Verhalten in Form des tatsächlichen Energieverbrauchs beobachtet werden kann. Akteure, die über bedeutende Expertise in diesem Gebiet verfügen, wie die Verbraucherzentrale Baden-Württemberg sowie die Klimaschutz- und Energieagentur Baden-Württemberg GmbH, sollten unterstützend in diese Projekte eingebunden werden. Zudem bietet es sich an, geplante oder laufende Aktivitäten der Verbraucherzentrale mit Blick auf Energiespareffekte begleitend zu evaluieren.

Für die Maßnahme Labeling gestaltet sich der Weg als umfassender, weil die gesetzlichen Vorgaben grundsätzlich auf EU-Ebene festgelegt werden. Die Verbraucherpolitik Baden-Württembergs kann hier einen zweistufigen Prozess anstoßen: Im ersten Schritt kann sie, wie für die anderen Maßnahmen, die Effekte in Feldexperimenten testen. Hier bieten sich insbesondere Forschungsprojekte zwischen Wissenschaft und Einzelhandel an. Im zweiten Schritt kann sie entweder versuchen, ihren Einfluss auf europäischer Ebene zu nutzen, um gesetzliche Vorgaben zu beeinflussen oder zu prüfen, ob zusätzlich zu den EU-Vorgaben ergänzende Informationen auf den Produkten sinnvoll und umsetzbar erscheinen. Frühere Empfehlungen der Verbraucherkommission Baden-Württemberg, z.B. diejenigen zu technischer Produktobsoleszenz (vgl. Brönneke et al. 2014), sollten hierbei berücksichtigt werden.

Nach einer systematischen Evaluation der Einspareffekte, einer Durchführung von Kosten-Nutzen- sowie Wohlfahrtsanalysen kann abschließend evidenzbasiert entschieden werden, ob und welche Maßnahmen wie umgesetzt werden. Eine solche Evaluation ist sinnvoll, damit dort Ressourcen investiert werden, wo sie den größten Nutzen stiften – in diesem Fall mit Blick auf das Ziel Energiesparen bzw. für den Klimaschutz.

6. Literatur

6.1. Zitierte Literatur

- Abrahamse, W., Steg, L., Vlek, C. & Rothengatter, T. (2005). A review of intervention studies aimed at household energy conservation. *Journal of Environmental Psychology*, 25(3), 273–291.
- acatech/Leopoldina/Akademienunion (2017). *Sektorkopplung - Optionen für die nächste Phase der Energiewende*. Schriftenreihe zur wissenschaftsbasierten Politikberatung.
- Allcott, H. (2011a). Rethinking real-time electricity pricing. *Resource and Energy Economics*, 33(4), 820–842.
- Allcott, H. (2015). Site selection bias in program evaluation. *The Quarterly Journal of Economics*, 130(3), 1117–1165.
- Allcott, H. (2016). Paternalism and energy efficiency: An overview. *Annual Review of Economics*, 8, 145–176.
- Allcott, H. & Kessler, J. B. (2017). *The welfare effects of nudges: A case study of energy use social comparisons*. *American Economic Journal: Applied Economics*, forthcoming.
- Allcott, H. & Mullainathan, S. (2010). Behavior and energy policy. *American Association for the Advancement of Science*, 327(5970), 1204-1205. Doi: 10.1126/science.1180775
- Andor, M. & Fels, K. (2017a). Behavioral economics and energy conservation – a systematic review of nonprice interventions and their causal effects. SFB 823, Discussion Paper Nr. 14/2017.
- Andor, M. & Fels, K. (2017b). Energiesparen durch verhaltensökonomisch motivierte Maßnahmen? acatech Schriftenreihe Energiesysteme der Zukunft.
- Andor, M., Gerster, A., Peters, J. & Schmidt, C. M. (2017). Social norms and energy conservation beyond the US. *Ruhr Economic Papers* #714.
- Angrist, J. D. & Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An empiricists' companion*. Princeton university press.
- Aste, N., Adhikari, R. S. & Manfren, M. (2013). Cost optimal analysis of heat pump technology adoption in residential reference buildings. *Renewable Energy*, 60, 615–624.
- Baldwin, R. (2014). From regulation to behaviour change: Giving nudge the third degree. *Modern Law Review*, 77, 831–857. Doi:10.1111/1468-2230.12094
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Upper Saddle River: Prentice Hall.
- Bauer, T. K., Fertig, M. & Schmidt, C. M. (2009). *Empirische Wirtschaftsforschung*. Berlin/Heidelberg.
- Bauwens, L., Fiebig, D. G.; & Steel, M. F. J. (1994). Estimating end-use demand: A Bayesian approach. *Journal of Business and Economic Statistics*, 12(2), 221–231.
- Becker, B. J. & Wu, M.-J. (2007). The synthesis of regression slopes in meta-analysis. *Statistical Science*, 22(3), 414–429. doi: 10.1214/07-STS 243.
- Bedwell, B., Leygue, C., Goulden, M., McAuley, D., Colley, J. & Ferguson, E. (2014). Apportioning energy consumption in the workplace: A review of issues in using metering data to motivate staff to save energy. *Technology Analysis and Strategic Management*, 26(10), 1196–1211.
- Benartzi, S., Beshears, J., Milkman, K. L., Sunstein, C. R., Thaler, R. H., Shankar, M., ... & Galing, S. (2017). Should governments invest more in nudging? *Psychological Science*, 28(8), 1041–1055.
- Boockmann, B., Buch, C. M. & Schnitzer, M. (2014). Evidenzbasierte Wirtschaftspolitik in Deutschland: Defizite und Potentiale. *Perspektiven der Wirtschaftspolitik*, 15(4), 307-323.
- Borenstein, M., Cooper, H., Hedges, L.V. & Valentine, J.C. (2009a). Effect sizes for continuous data. In: Cooper, H., Hedges, V. & Valentine, J. C. (Hsg.), *The Handbook of Research Synthesis and Meta-analysis* (S. 221–235). New York: Russell Sage Foundation.

- Borenstein M., Hedges, L. V., Higgins, J. P. T. & Rothstein, H. R. (2009b). Introduction to meta-analysis. Chichester, UK: John Wiley and Sons.
- Bosworth, S. & Bartke, S. (2014). Implikationen von Nudging für das Wohlergehen von Konsumenten. Beitrag zum Zeitgespräch. Wirtschaftsdienst 11, November 2014.
- Bovens, L. (2009). The ethics of nudge. In: Grüne-Yanoff, T., Hansson, S. O. (Hrs.). *Preference Change: Approaches from philosophy, economics and psychology*. Theory and Decision Library, Band 42. Dodrecht: Springer.
- Brönneke, T., Fezer, K.-H., Oehler, A., Reisch, L. & Stellpflug, J. (2014). *Qualität statt vorzeitiger Verschleiß – Diskussionspapier zur eingebauten Obsoleszenz bei Konsumgütern*. Stellungnahme Verbraucherkommission Baden-Württemberg Nr. 32a/2014.
- Brutscher, P.-B. (2011). *Payment matters? – An exploratory study into the pre-payment electricity metering* (Cambridge Working Paper in Economics No. 1124). University of Cambridge, Cambridge. Doi: 10.17863/CAM.5525
- Buchanan, K., Russo, R. & Anderson, B. (2014). Feeding back about eco-feedback: How do consumers use and respond to energy monitors? *Energy Policy*, 73, 138–146.
- Buryk, S., Mead, D., Mourato, S. & Torriti, J. (2015). Investigating preferences for dynamic electricity tariffs: The effect of environmental and system benefit disclosure. *Energy Policy*, 80, 190–195.
- Campbell Collaboration (2014). Campbell Systematic Reviews: Policies and Guidelines. Version 1.2. Campbell Policies and Guidelines Series No. 1.
- Chetty, R. (2015). Behavioral economics and public policy: A pragmatic perspective. *The American Economic Review*, 105(5), 1–33.
- Ciriolo, E. (2011). Behavioural economics in the European Commission: Past, present and future. Oxera, Agenda, Jan. 2011.
- Clee, M. A. & Wicklund, R. A. (1980). Consumer behavior and psychological reactance. *Journal of Consumer Research*, 6(4), 389–405.
- Cohn, L. D. & Becker, B.J. (2003). How meta-analysis increases statistical power. *Psychological methods* 8 (3), 243–253.
- Damgaard, M. T. & Gravert, C. (2018). The hidden costs of nudging: Experimental evidence from reminders in fundraising *Journal of Public Economics*, 157, 15–26, forthcoming.
- Darby, S. (2006). The effectiveness of feedback on energy consumption. A Review for DEFRA of the Literature on Metering, Billing and direct Displays, Oxford: Oxford Environmental Change Institute.
- Delmas, M. A., Fischlein, M. & Asensio, O. I. (2013). Information strategies and energy conservation behavior: A meta-analysis of experimental studies from 1975 to 2012. *Energy Policy*, 61, 729–739.
- DerSimonian, R. & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7(3), 177–188.
- Di Cosmo, V., Lyons, S. & Nolan, A. (2014). Estimating the impact of time-of-use pricing on Irish electricity demand. *Energy Journal*, 35(2), 117–136.
- Dogan, E., Bolderdijk, J. W. & Steg, L. (2014). Making small numbers count: Environmental and financial feedback in promoting eco-driving behaviours. *Journal of Consumer Policy*, 37(3), 413–422.
- Dowd, A., Ashworth, P., Carr-Cornish, S. & Stenner, K. (2012). Energymark: Empowering individual Australians to reduce their energy consumption. *Energy Policy*, 51(1), 264–276.
- Ehrhardt-Martinez, K., Donnelly, K. A. & Laitner, S. (2010). *Advanced metering initiatives and residential feedback programs: A meta-review for household electricity-saving opportunities*. Washington, DC: American Council for an Energy-Efficient Economy.
- Ernst, A., Reisch, L., Renn, O., Andor, M., Gallego, D., Münch, W., Paschke, M. & Schetula, M. (2017, i.E.) *Verbraucherpolitik für die Energiewende*. acatech, Leopoldina und Akademiunion (Hrs.). Schriftenreihe

zur wissenschaftsbasierten Politikberatung, München.

Fischer, C. (2008). Feedback on household electricity consumption: A tool for saving energy? *Energy efficiency*, 1(1), 79–104.

Fronzel, M. & Vance, C. (2013). Energy efficiency: Don't belittle the rebound effect. *Nature*, 494, 430–430. Doi: 10.1038/494430c

Gigerenzer, G. (2015). On the supposed evidence for libertarian paternalism. *Review of Philosophy and Psychology*, 6(3), 361–383.

Glass, G. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, 5, 3–8.

Greenhalgh, T. (2014). *How to read a paper: The basics of evidence-based medicine*. London: BMJ Books.

Hansen, P. & Jespersen, A. (2013). Nudge and the manipulation of choice: A framework for the responsible use of the nudge approach to behaviour change in public policy. *European Journal of Risk Regulation*, 4(1), 3–28. Doi:10.1017/S1867299X00002762

Hargreaves, T., Nye, M. & Burgess, J. (2010). Making energy visible: A qualitative field study of how householders interact with feedback from smart energy monitors. *Energy Policy*, 38(10), 6111–6119.

Hargreaves, T., Nye, M. & Burgess, J. (2013). Keeping energy visible? Exploring how householders interact with feedback from smart energy monitors in the longer term. *Energy Policy*, 52(1), 126–134.

Heberlein, T. A. & Warriner, G. K. (1983). The influence of price and attitude on shifting residential electricity consumption from on- to off-peak periods. *Journal of Economic Psychology*, 4(1–2), 107–130.

Hedges, L. V., Gurevitch, J. & Curtis, P.S. (1999). The meta-analysis of response ratios in experimental ecology. *Ecology*, 80(4): 1150–1156.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 6, 107–128.

Higgins, J. P. T. & Green, S. (2011). *Cochrane handbook for systematic review of interventions*. Version 5.1.0 [updated March 2011]. The Cochrane Collaboration. Verfügbar unter: www.cochrane-handbook.org.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J. & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ: British Medical Journal*, 327(7414), 557–560.

Higgins, J. P. T., White, I. R. & Anzueto-Cabrera J. (2008). Meta-analysis of skewed data: Combining results reported on log-transformed or raw scales. *Statistics in Medicine*, 27, 6072–6092. Doi: 10.1002/sim.3427.

Hille, S. L. (2016). The myth of the unscrupulous energy user's dilemma: Evidence from Switzerland. *Journal of Consumer Policy*, 39(3), 327–347.

Houle, D., Pelabon, C., Wagner, G. P. & Hansen, T. F. (2011). Measurement and meaning in biology. *Quarterly Review of Biology*, 86, 3–34.

House of Lords. (2011). *Behaviour Change: 2nd report of session 2010-12 of the Science and Technology Committee*. Verfügbar unter: <http://www.publications.parliament.uk/pa/ld201012/ldselect/ldsctech/179/179.pdf>.

Hunter, J. E. & Schmidt, F. L. (2000). Fixed effects vs. random effects in meta-analysis models: Implications for cumulative research knowledge. *International Journal of Selection and Assessment*, 8, 275–292. Doi: 10.1111/1468-2389.00156.

IAWG (2013). *Technical support document: Technical update of the social cost of carbon for regulatory impact analysis*. U.S. Interagency Working Group on the Social Cost of Carbon for Regulatory Impact Analysis. Verfügbar unter: https://www.whitehouse.gov/sites/default/files/omb/inforeg/social_cost_of_carbon_for_ria_2013_update.pdf.

Imbens, G. W. & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47(1), 5–86.

- Ivanov, C., Getachew, L., Fenrick, S. A. & Vittetoe, B. (2013). Enabling technologies and energy savings: The case of energywise smart meter pilot of Connexus Energy. *Utilities Policy*, 26, 76–84.
- Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioural economics. *The American Economic Review*, 93(5), 1449–1475.
- Karlin, B., Ford, R. & Zinger, J. (2015). The effects of feedback on energy conservation: A meta-analysis. *Psychological Bulletin*, 141, 1205–1247.
- Kasperbauer, T. J. (2017). The permissibility of nudging for sustainable energy consumption. *Energy Policy*, 111, 52–57.
- Keef, S.P. & Roberts, L. A. (2004). The meta-analysis of partial effect sizes. *The British Journal of Mathematical and Statistical Psychology*, 57(1), 97–129. Doi: 10.1348/000711004849303.
- Kelley, G. A. & Kelley, K. S. (2012). Statistical models for meta-analysis: A brief tutorial. *World Journal of Methodology*, 2(4), 27–32.
- Kirchgässner, G. (2017). Soft paternalism, merit goods, and normative individualism. *European Journal of Law and Economics*, 43(1), 125–152.
- Komatsu, H. & Nishio, K. (2015). An experimental study on motivational change for electricity conservation by normative messages. *Applied Energy* 158, 35–43. DOI: 10.1016/j.apenergy.2015.08.029.
- Köszegi, B. & Szeidl, A. (2012). A model of focusing in economic choice. *The Quarterly Journal of Economics*, 128(1), 53–104.
- Laibson, D. (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics*, 112(2), 443–478.
- Lang, C. & Okwelum, E. (2015). The mitigating effect of strategic behavior on the net benefits of a direct load control program. *Energy Economics*, 49, 141–148.
- Litvine, D. & Wustenhagen, R. (2011). Helping ‘Light Green’ Consumers Walk the Talk: Results of a Behavioural Intervention Survey in the Swiss Electricity Market. *Ecological Economics*, 70(3), 462–474.
- Lokhorst, A. M., Werner, C., Staats, H., van Dijk, E. & Gale, J. E. (2013). Commitment and behavior change: A meta-analysis and critical review of commitment-making strategies in environmental research. *Environment and Behavior*, 45(1), 3–34.
- Lutzenhiser, L. (1992). A cultural model of household energy consumption. *Energy*, 17(1), 47–60.
- Madrian, B. C. (2014). Applying insights from behavioral economics to policy design. *Annual Review of Economics*, 6(1), 663–688.
- Mills, B. & Schleich, J. (2010). What’s driving energy efficient appliance label awareness and purchase propensity? *Energy Policy*, 38(2), 814–825.
- Murtagh, N. (2013). Individual energy use and feedback in an office setting: A field trial. *Energy Policy*, 62, 717–728.
- Nakagawa, S., Noble, D. W., Senior, A. M. & Lagisz, M. (2017). Meta-evaluation of meta-analysis: Ten appraisal questions for biologists. *BMC biology*, 15(1), 1–14.
- Nisan, M. & Horenczyk, G. (1990). Moral balance: The effect of prior behaviour on decision in moral conflict. *British Journal of Social Psychology*, 29(1), 29–42.
- Nordhaus, W. (2014). Estimates of the social cost of carbon: Concepts and results from the DICE-2013R model and alternative approaches. *Journal of the Association of Environmental and Resource Economists* 1 (1/2), 273–312.
- O’Donoghue, T. & Rabin, M. (1999). Doing it now or later. *The American Economic Review*, 89(1), 103–124.
- O’Donoghue, T. & Rabin, M. (2008). Procrastination on long-term projects. *Journal of Economic Behavior & Organization*, 66(2), 161–175.
- Oltra, C., Boso, A., Espluga, J. & Prades, A. (2013). A qualitative study of users’ engagement with real-time

- feedback from in-house energy consumption displays. *Energy Policy*, 61, 788–792.
- Oseni, M. O., Pollitt, M. G., Reiner, D. M., Richter, L. L., Chyong, K. & Baddeley, M. (2013). 2013 EPRG Public Opinion Survey: Smart Energy Survey-Attitudes and Behaviours (Cambridge Working Papers in Economics No. 1352), University of Cambridge, Cambridge.
- Paetz, A.-G., Dutschke, E. & Fichtner, W. (2012). Smart homes as a means to sustainable energy consumption: A study of consumer perceptions. *Journal of Consumer Policy*, 35(1), 23–41.
- Partlett, C. & Riley, R. D. (2017). Random effects meta-analysis: Coverage performance of 95% confidence and prediction intervals following REML estimation. *Statistics in medicine*, 36(2), 301–317.
- Pigott, T. D. (2009). Handling missing data. In: Cooper, H. M., Hedges, L. V. & Valentine, J. C. (Hrs.), *The handbook of research synthesis and meta-analysis* (S. 399–416). New York, NY: Russell Sage Foundation.
- Pullinger, M., Lovell, H. & Webb, J. (2014). Influencing household energy practices: A critical review of UK smart metering standards and commercial feedback devices. *Technology Analysis and Strategic Management*, 26(10), 1144–1162.
- Purnhagen, Kai P. & Reisch, L. A. (2016). 'Nudging Germany'? Herausforderungen für eine verhaltensbasierte Regulierung in Deutschland. *Zeitschrift für Europäisches Privatrecht*, 3/2016, 629–655
- Reisch, L. A. (2013b). *Elemente einer verhaltensorientierten Energienachfragepolitik*. In: Held, M., Kubon-Gilke, G. & Sturn, R. (Hrs.), *Grenzen der Konsumentensouveränität* (S. 139–159). Marburg: Metropolis-Verlag. (Jahrbuch Normative und institutionelle Grundfragen der Ökonomik, Vol. 12).
- Reisch, L. A. & Sunstein, C. R. (2016). Do Europeans like nudges? *Judgment and Decision Making*, 11(4), 310–325.
- Reisch, L. A. & Sunstein, C. R. (2017). Verhaltensbasierte Regulierung (Nudging). In: Kenning, P., Oehler, A., Reisch, L. A. & Grugel, C. (Hrs.), *Verbraucherwissenschaften* (S. 341–365). Wiesbaden: Springer Gabler.
- Richalet, V., Neirac, F. P., Tellez, F., Marco, J. & Bloem, J. J. (2001). HELP (house energy labeling procedure): methodology and present results. *Energy and Buildings*, 33(3), 229–233.
- Roberts, S. & Baker, W. (2003). Towards effective energy information. Improving consumer feedback on energy consumption. A report to OFGEM. Verfügbar unter: https://www.ofgem.gov.uk/sites/default/files/docs/2003/07/4138-towards_effective_energy_information_cse_report_ofgem_july03.pdf.
- Rohling, M. & Schubert, R. (2013). Energy labels for household appliances and their disclosure format: a literature review. *IED Working Paper*, 21. Zürich: ETH Zürich.
- Rosenthal, R. & DiMatteo, M. R. (2001). Meta-analysis: Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology*, 52(1), 59–82.
- RWI (2017). *Energiesparen in privaten Haushalten – Ein Randomized Controlled Trial zur Wirkungsevaluierung einer Endverbrauchersensibilisierung*. Projektbericht für ein Forschungsprojekt gefördert durch die Stiftung Mercator.
- Schmidt, C. M. (2014). Wirkungstreffer erzielen – Die Rolle der evidenzbasierten Politikberatung in einer aufgeklärten Gesellschaft. *Perspektiven der Wirtschaftspolitik*, 15(3), 219–233.
- Schubert, R. & Stadelmann, M. (2015). Energy-using durables – why consumers refrain from economically optimal choices. *Frontiers in Energy Research*, 3(7), 1–13. Doi:10.3389/fenrg.2015.00007
- Shen, M., Cui, Q. & Fu, L. (2015). Personality traits and energy conservation. *Energy Policy*, 85, 322–334.
- Sorrell, S. (2015). Reducing energy demand: A review of issues, challenges and approaches. *Renewable and Sustainable Energy Reviews*, 47, 74–82.
- Stanley, T. D., Doucouliagos, H., Giles, M., Heckemeyer, J. H., Johnston, R. J., Laroche, P., ... & Rosenberger, R. S. (2013). Meta-analysis of economics research reporting guidelines. *Journal of Economic Surveys*, 27(2), 390–394.
- Steg, L. (2008). Promoting household energy conservation. *At the crossroads: Pathways of renewable and*

nuclear energy policy in North Africa, 36(12), 4449–4453.

Sterne, J. A., Sutton, A. J., Ioannidis, J. P., Terrin, N., Jones, D. R., Lau, J., ... & Tetzlaff, J. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *British Medical Journal*, 343, d4002.

Stillwater, T. & Kurani, K. S. (2013). Drivers discuss ecodriving feedback: Goal setting, framing, and anchoring motivate new behaviors. *Transportation Research Part F: Traffic Psychology and Behaviour*, 19, 85–96.

Sunstein, C. R. (2015). The ethics of nudging. *Yale Journal on Regulation*, 32(2), 413–450.

Sunstein, C. R. (2014). Nudging: A very short guide. *Journal of Consumer Policy*, 37(4), 583–588.

Sunstein, C. R. (2017a). Do people like nudges? *Administrative Law Review*, forthcoming.

Sunstein, C. R. (2017b). Misconceptions about nudges. Working Paper. Verfügbar unter: <https://ssrn.com/abstract=3033101>

Sunstein, C. R. & Reisch, L. A. (2014). Automatically green: Behavioral economics and environmental protection. *Harvard Environmental Law Review*, 38(1), 127–158.

Sunstein, C. R., Reisch, L. A. & Rauber, J. (2017). A world-wide consensus on nudging? Not quite, but almost. *Regulation and Governance*, forthcoming. DOI:10.1111/rego.12161.

Thaler, R. H. & Benartzi, S. (2004). Save more tomorrow: Using behavioral economics to increase employee saving. *Journal of Political Economy*, 112, 164–187.

Thaler, R. H. & Sunstein, C. R. (2003). Libertarian paternalism. *The American Economic Review*, 93(2), 175–179.

Thaler, R.H. & Sunstein, C. R. (2008). *Nudge. Improving decisions about health, wealth and happiness*. London: Penguin.

Thaler, R. H. & Sunstein, C. R. (2009). *Nudge: Wie man kluge Entscheidungen anstößt*. Berlin: Econ Verlag.

Vellei, M., Natarajan, S., Biri, B., Padget, J. & Walker, I. (2016). The effect of real-time context-aware feedback on occupants' heating behaviour and thermal adaptation. *Energy and Buildings*, 123, 179–191. Doi:10.1016/j.enbuild.2016.03.045

Veroniki, A. A., Jackson, D., Viechtbauer, W., Bender, R., Bowden, J., Knapp, G., ... & Salanti, G. (2016). Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Research Synthesis Methods*, 7(1), 55–79.

Vooren, M., Haelermans, C., Groot, W. & van den Brink, H. M. (2017). *The effectiveness of active labor market policies: A meta-analysis* (TIER Working Paper Series WP 17/01). Maastricht University, Maastricht.

Wang, M., Che, Y., Yang, K., Wang, M., Xiong, L. & Huang, Y. (2011). A local-scale low-carbon plan based on the STIRPAT model and the scenario method: The Case of Minhang District, Shanghai, China. *Energy Policy*, 39(11), 6981–6990.

Weckmann, G., Chenot, J. F. & Reber, K. C. (2015). Metaanalysen lesen und interpretieren: eine praktische Anleitung. *Zeitschrift für Allgemeinmedizin*, 91(11), 469–473. Doi 10.3238/zfa.2015.0469–0473

World Bank (2017). World Bank Country and Lending Groups. Historical classification by income in XLS format. Verfügbar unter: <http://databank.worldbank.org/data/download/site-content/OGHIST.xls>.

Yeomans, M. & Herberich, D. (2014). An experimental test of the effect of negative social norms on energy-efficient investments. *Journal of Economic Behavior and Organization*, 108, 187–197.

Zhou, N., Fridley, D., McNeil, M., Zheng, N., Letschert, V., Ke, J. & Saheb, Y. (2011). Analysis of potential energy saving and CO2 emission reduction of home appliances and commercial equipments in China. *At the Crossroads: Pathways of Renewable and Nuclear Energy Policy in North Africa*, 39(8), 4541–4550.

6.2. In die Meta-Analyse einbezogene Literatur

- Abrahamse, W., Steg, L., Vlek, C. & Rothengatter, T. (2007). The effect of tailored information, goal setting, and tailored feedback on household energy use, energy-related behaviors, and behavioral antecedents. *Journal of Environmental Psychology, 27*(4), 265–276. Doi:10.1016/j.jenvp.2007.08.002
- Alahmad, M. A., Wheeler, P. G., Schwer, A., Eiden, J. & Brumbaugh, A. (2012). A comparative study of three feedback devices for residential real-time energy monitoring. *IEEE Transactions on Industrial Electronics, 59*(4), 2002–2013. Doi:10.1109/TIE.2011.2165456
- Alberts, G., Gurguc, Z., Koutroumpis, P., Martin, R., Muuls, M. & Napp, T. (2016). Competition and norms: a self-defeating combination? *Energy Policy, 96*, 504–523. Doi:10.1016/j.enpol.2016.06.001
- Allcott, H. (2011b). Social norms and energy conservation. *Journal of Public Economics, 95*(9-10), 1082–1095. Doi:10.1016/j.jpubeco.2011.03.003
- Allcott, H. & Rogers, T. (2014). The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. *American Economic Review, 104*(10), 3003–3037. Doi:10.1257/aer.104.10.3003
- Allcott, H. & Taubinsky, D. (2015). Evaluating behaviorally motivated policy: Experimental evidence from the lightbulb market. *American Economic Review, 105*(8), 2501–2538. Doi:10.1257/aer.20131564
- Allen, D. & Janda, K. (2006). The effects of household characteristics and energy use consciousness on the effectiveness of real-time energy use feedback: A pilot study. *Proceedings of ACEEE Summer Study*, 1–12.
- Anderson, C. D. & Claxton, J. D. (1982). Barriers to consumer choice of energy efficient products. *Journal of Consumer Research, 9*(2), 163–170. Retrieved from <http://www.jstor.org/stable/2489125>
- Anderson, K., Song, K., Lee, S., Krupka, E., Lee, H. & Park, M. (2017). Longitudinal analysis of normative energy use feedback on dormitory occupants. *Applied Energy, 189*, 623–639. Doi:10.1016/j.apenergy.2016.12.086
- Arvola, A., Uutela, A. & Anttila, U. (1993). Billing feedback as means to encourage household electricity conservation: A field experiment in Helsinki. *Proceedings of eceee Summer Study*, 11–21.
- Ayres, I., Raseman, S. & Shih, A. (2013). Evidence from two large field experiments that peer comparison feedback can reduce residential energy usage. *Journal of Law, Economics, and Organization, 29*(5), 992–1022. Doi:10.1093/jleo/ews020
- Battalio, R. C., Kagel, J. H. & Winkler, R. C. (1979). Residential electricity demand: An experimental study. *The Review of Economics and Statistics, 61*(2), 180–189. Doi:10.2307/1924585
- Becker, L. J. (1978). Joint effect of feedback and goal setting on performance: a field study of residential energy conservation. *Journal of Applied Psychology, 63*(4), 428–433. Doi:10.1037/0021-9010.63.4.428
- Bekker, M. J., Cumming, T. D., Osborne, N. K. P., Bruining, A. M., McClean, J. I. & Leland, L. S. (2010). Encouraging electricity savings in a university residential hall through a combination of feedback, visual prompts, and incentives. *Journal of Applied Behavior Analysis, 43*(2), 327–331. Doi:10.1901/jaba.2010.43-327
- Benders, R. M., Kok, R., Moll, H. C., Wiersma, G. & Noorman, K. J. (2006). New approaches for household energy conservation—In search of personal household energy budgets and energy reduction options. *Energy Policy, 34*(18), 3612–3622. Doi:10.1016/j.enpol.2005.08.005
- Bittle, R. G., Valesano, R. M. & Thaler, G. M. (1979a). The effects of daily feedback on residential electricity usage as a function of usage level and type of feedback information. *Journal of Environmental Systems, 9*(3), 275–287. Doi:10.2190/91AA-P97G-JF92-T7EJ
- Bittle, R. G., Valesano, R. M. & Thaler, G. M. (1979b). The effects of daily cost feedback on residential electricity consumption. *Behavior Modification, 3*(2), 187–202. Doi:10.1177/014544557932004
- Brandon, G. & Lewis, A. (1999). Reducing household energy consumption: A qualitative and quantitative field study. *Journal of Environmental Psychology, 19*(1), 75–85. Doi:10.1006/jevp.1998.0105

- Brent, D. A., Cook, J. H. & Olsen, S. (2015). Social comparisons, household water use, and participation in utility conservation programs: Evidence from three randomized trials. *Journal of the Association of Environmental and Resource Economists*, 2(4), 597–627. Doi:10.1086/683427
- Brounen, D. & Kok, N. (2011). On the economics of energy labels in the housing market. *Journal of Environmental Economics and Management*, 62(2), 166–179. Doi:10.1016/j.jeem.2010.11.006
- Carroll, J., Lyons, S. & Denny, E. (2014). Reducing household electricity demand through smart metering: The role of improved information about energy saving. *Energy Economics*, 45, 234–243. Doi:10.1016/j.eneco.2014.07.007
- Costa, D. L. & Kahn, M. E. (2013). Energy conservation ‘nudges’ and environmentalist ideology: Evidence from a randomized residential electricity field experiment. *Journal of the European Economic Association*, 11(3), 680–702. Doi:10.1111/jeea.12011
- Delmas, M. A. & Lessem, N. (2014). Saving power to conserve your reputation? The effectiveness of private versus public information. *Journal of Environmental Economics and Management*, 67(3), 353–370. Doi:10.1016/j.jeem.2013.12.009
- Dobson, J. K. & Griffin, J. A. (1992). Conservation effect of immediate electricity cost feedback on residential consumption behaviour. *Proceedings of ACEEE Summer Study*, 33–35.
- Dolan, P. & Metcalfe, R. (2015). *Neighbors, knowledge, and nuggets: Two natural field experiments on the role of incentives on energy conservation* (Becker Friedman Institute for Research in Economics Working Paper No. 2589269). Becker Friedman Institute for Research, Chicago.
- Emeakaroha, A., Ang, C. S., Yan, Y. & Hopthrow, T. (2014). Integrating persuasive technology with energy delegates for energy conservation and carbon emission reduction in a university campus. *Energy*, 76, 357–374. Doi:10.1016/j.energy.2014.08.027
- Ferraro, P. J., Miranda, J. J. & Price, M. K. (2011). The persistence of treatment effects with norm-based policy instruments: Evidence from a randomized environmental policy experiment. *American Economic Review*, 101(3), 318–322. Doi:10.1257/aer.101.3.318
- Ferraro, P. J. & Price, M. K. (2013). Using nonpecuniary strategies to influence behavior: Evidence from a large-scale field experiment. *Review of Economics and Statistics*, 95(1), 64–73. Doi:10.1162/REST_a_00344
- Fielding, K. S., Spinks, A., Russell, S., McCrea, R., Stewart, R. & Gardner, J. (2013). An experimental test of voluntary strategies to promote urban water demand management. *Journal of Environmental Management*, 114, 343–351. Doi:10.1016/j.jenvman.2012.10.027
- Gans, W., Alberini, A. & Longo, A. (2013). Smart meter devices and the effect of feedback on residential electricity consumption: Evidence from a natural experiment in Northern Ireland. *Energy Economics*, 36, 729–743. Doi:10.1016/j.eneco.2012.11.022
- Geller, E. S., Erickson, J. B. & Buttram, B. A. (1983). Attempts to promote residential water conservation with educational, behavioral and engineering strategies. *Population and Environment*, 6(2), 96–112.
- Gilbert, B. & Graff Zivin, J. (2014). Dynamic salience with intermittent billing: Evidence from smart electricity meters. *Journal of Economic Behavior and Organization*, 107, 176–190. Doi:10.1016/j.jebo.2014.03.011
- Gleerup, M., Larsen, A., Leth-Petersen, S. & Togeby, M. (2010). The Effect of feedback by text message (SMS) and email on household electricity consumption: Experimental evidence. *Energy Journal*, 31(3), 113–132.
- Grønhoj, A. & Thøgersen, J. (2011). Feedback on household electricity consumption: Learning and social influence processes. *International Journal of Consumer Studies*, 35(2), 138–145. Doi:10.1111/j.1470-6431.2010.00967.x
- Haakana, M., Sillanpää, L. & Talsi, M. (1997). The effect of feedback and focused advice on household energy consumption. *Proceedings of ecee Summer Study*, 1–11.
- Hahn, R., Metcalfe, R. D., Novgorodsky, D. & Price, M. K. (2016). *The behavioralist as policy designer: The*

- need to test multiple treatments to meet multiple targets (NBER Working Paper No. 22886). National Bureau of Economic Research, Cambridge.
- Harding, M. & Hsiaw, A. (2014). Goal setting and energy conservation. *Journal of Economic Behavior and Organization*, 107, 209–227. Doi:10.1016/j.jebo.2014.04.012
- Hayes, S. C. & Cone, J. D. (1981). Reduction of residential consumption of electricity through simple monthly feedback. *Journal of Applied Behavior Analysis*, 14(1), 81–88. Doi:10.1901/jaba.1981.14-81
- Heinzle, S. L. (2012). Disclosure of energy operating cost information: A silver bullet for overcoming the energy-efficiency gap? *Journal of Consumer Policy*, 35(1), 43–64. Doi:10.1007/s10603-012-9189-6
- Houde, S. (2014). *How consumers respond to environmental certification and the value of energy information* (NBER Working Paper No. 20019). National Bureau of Economic Research, Cambridge.
- Houde, S., Todd, A., Sudarshan, A., Flora, J. A. & Armel, K. Carrie. (2013). Real-time feedback and electricity consumption: A field experiment assessing the potential for savings and persistence. *Energy Journal*, 34(1), 87–102. Doi:10.5547/01956574.34.1.4
- Hutton, R. B. & Mauser, G. A. (1986). Effects of cost-related feedback on consumer knowledge and consumption behavior: A field experimental approach. *Journal of Consumer Research*, 13(3), 327–336. Verfügbar unter: <http://www.jstor.org/stable/2489424>
- Ivanov, C., Getachew, L., Fenrick, S. A. & Vittetoe, B. (2013). Enabling technologies and energy savings: The case of EnergyWise Smart Meter Pilot of Connexus Energy. *Utilities Policy*, 26, 76–84. Doi:10.1016/j.jup.2012.10.001
- Jaeger, C. M. & Schultz, P. Wesley. (2017). Coupling social norms and commitments: Testing the underdetected nature of social influence. *Journal of Environmental Psychology*, 51, 199–208. Doi:10.1016/j.jenvp.2017.03.015
- Jain, R. K., Taylor, J. E. & Culligan, P. J. (2013). Investigating the impact eco-feedback information representation has on building occupant energy consumption behavior and savings. *Energy and Buildings*, 64, 408–414. Doi:10.1016/j.enbuild.2013.05.011
- Jensen, O. M., Hansen, A. R. & Kragh, J. (2016). Market response to the public display of energy performance rating at property sales. *Energy Policy*, 93, 229–235. Doi:10.1016/j.enpol.2016.02.029
- Jeong, S. H., Gulbinas, R., Jain, R. K. & Taylor, J. E. (2014). The impact of combined water and energy consumption eco-feedback on conservation. *Energy and Buildings*, 80, 114–119. Doi:10.1016/j.enbuild.2014.05.013
- Kantola, S. J., Syme, G. J. & Campbell, N. A. (1984). Cognitive dissonance and energy conservation. *Journal of Applied Psychology*, 69(3), 416–421. Doi:10.1037//0021-9010.69.3.416
- Katzev, R., Cooper, L. & Fisher, P. (1980). The effect of feedback and social reinforcement on residential electricity consumption. *Journal of Environmental Systems*, 10(3), 215–227. Doi:10.2190/M8W4-JUYV-NB03-CUFL
- Katzev, R. & Johnson, T. (1983). A social-psychological analysis of residential electricity consumption: The impact of minimal justification techniques. *Journal of Economic Psychology*, 3(3-4), 267–284. Doi:10.1016/0167-4870(83)90006-5
- Katzev, R. & Johnson, T. (1984). Comparing the effects of monetary incentives and foot-in-the-door strategies in promoting residential electricity conservation. *Journal of Applied Social Psychology*, 14(1), 12–27. Doi:10.1111/j.1559-1816.1984.tb02217.x
- Kua, H. W. & Wong, S. E. (2012). Lessons for integrated household energy conservation policies from an intervention study in Singapore. *Energy Policy*, 47(1), 49–56. Doi:10.1016/j.enpol.2012.04.009
- Kurz, T., Donaghue, N. & Walker, I. (2005). Utilizing a social-ecological framework to promote water and energy conservation: A field experiment. *Journal of Applied Social Psychology*, 35(6), 1281–1300. Doi:10.1111/j.1559-1816.2005.tb02171.x
- Lynham, J., Nitta, K., Saijo, T. & Tarui, N. (2016). Why does real-time information reduce energy

- consumption? *Energy Economics*, 54, 173–181. Doi:10.1016/j.eneco.2015.11.007
- Matsukawa, I. (2004). The effects of information on residential demand for electricity. *Energy Journal*, 25(1), 1–17. Retrieved from <http://www.jstor.org/stable/41323018>
- McCalley, L. T. & Midden, Cees J. H. (2002). Energy conservation through product-integrated feedback: The roles of goal-setting and social orientation. *Journal of Economic Psychology*, 23(5), 589–603. Doi:10.1016/S0167-4870(02)00119-8
- McClelland, L. & Cook, S. W. (1979). Energy conservation effects of continuous in-home feedback in all-electric homes. *Journal of Environmental Systems*, 9(2), 169–173. Doi:10.2190/L8BU-ECLK-PEC5-KKTW
- Midden, C. J., Meter, J. F., Weenig, M. H. & Zieverink, H. J. (1983). Using feedback, reinforcement and information to reduce energy consumption in households: A field-experiment. *Journal of Economic Psychology*, 3(1), 65–86. Doi:10.1016/0167-4870(83)90058-2
- Mizobuchi, K. & Takeuchi, K. (2013). The influences of financial and non-financial factors on energy-saving behaviour: A field experiment in Japan. *Energy Policy*, 63, 775–787. Doi:10.1016/j.enpol.2013.08.064
- Mountain, D. C. (2012). *Real-time feedback and residential electricity consumption: The Newfoundland and Labrador pilot* (QSEP Research Report No. 449). McMaster University, Hamilton.
- Newell, R. G. & Siikamäki, J. V. (2014). Nudging energy efficiency behavior: The role of information labels. *Journal of the Association of Environmental and Resource Economists*, 1(4), 555–598. Doi:10.1086/679281
- Nilsson, A., Bergstad, C. J., Thuvander, L., Andersson, D., Andersson, K. & Meiling, P. (2014). Effects of continuous feedback on households' electricity consumption: Potentials and barriers. *Applied Energy*, 122(1), 17–23. Doi:10.1016/j.apenergy.2014.01.060
- Ölander, F. & Thøgersen, J. (2014). Informing versus nudging in environmental policy. *Journal of Consumer Policy*, 37(3), 341–356. Doi:10.1007/s10603-014-9256-2
- Peschiera, G. & Taylor, J. E. (2012). The impact of peer network position on electricity consumption in building occupant networks utilizing energy feedback systems. *Energy and Buildings*, 49, 584–590. Doi:10.1016/j.enbuild.2012.03.011
- Peschiera, G., Taylor, J. E. & Siegel, J. A. (2010). Response–relapse patterns of building occupant electricity consumption following exposure to personal, contextualized and occupant peer network utilization data. *Energy and Buildings*, 42(8), 1329–1336. Doi:10.1016/j.enbuild.2010.03.001
- Schleich, J., Klobasa, M., Golz, S. & Brunner, M. (2013). Effects of feedback on residential electricity demand—findings from a field trial in Austria. *Energy Policy*, 61, 1097–1106. Doi:10.1016/j.enpol.2013.05.012
- Schultz, P. Wesley, Estrada, M., Schmitt, J., Sokoloski, R. & Silva-Send, N. (2015). Using in-home displays to provide smart meter feedback about household electricity consumption: A randomized control trial comparing kilowatts, cost, and social norms. *Energy*, 90(Part 1), 351–358. Doi:10.1016/j.energy.2015.06.130
- Schultz, P. Wesley, Nolan, J. M., Cialdini, R. B., Goldstein, N. J. & Giskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological Science*, 18(5), 429–434. Doi:10.1111/j.1467-9280.2007.01917.x
- Seligman, C., Darley, J. M. & Becker, L. J. (1978). Behavioral approaches to residential energy conservation. *Energy and Buildings*, 1(3), 325–337. Doi:10.1016/0378-7788(78)90012-9
- Seyranian, V., Sinatra, G. M. & Polikoff, M. S. (2015). Comparing communication strategies for reducing residential water consumption. *Journal of Environmental Psychology*, 41, 81–90. Doi:10.1016/j.jenvp.2014.11.009
- Shen, M., Young, R. & Cui, Q. (2016). The normative feedback approach for energy conservation behavior in the military community. *Energy Policy*, 98, 19–32. Doi:10.1016/j.enpol.2016.08.014
- Sipe, B. & Castor, S. (2009). *The net impact of home energy feedback devices*. Proceedings from IEPEC '09: International Energy Program Evaluation Conference, Portland: IEPEC.

- Staats, H., Harland, P. & Wilke, H. A. M. (2004). Effecting durable change: A team approach to improve environmental behavior in the household. *Environment and Behavior*, 36(3), 341–367. Doi:10.1177/0013916503260163
- Stinson, J., Willis, A., Williamson, J. B., Currie, J. & Smith, R. (2015). Visualising energy use for smart homes and informed users. *Energy Procedia*, 78, 579–584. Doi:10.1016/j.egypro.2015.11.015
- Tiefenbeck, V., Goette, L., Degen, K., Tasic, V., Fleisch, E., Lalive, R. & Staake, T. (2017). Overcoming salience bias: How real-time feedback fosters resource conservation. *Management Science*, forthcoming.
- Tiefenbeck, V., Staake, T., Roth, K. & Sachs, O. (2013). For better or for worse? Empirical evidence of moral licensing in a behavioral energy conservation campaign. *Energy Policy*, 57, 160–171. Doi:10.1016/j.enpol.2013.01.021
- van Dam, S. S., Bakker, C. A. & van Hal, J. D. M. (2010). Home energy monitors: Impact over the medium-term. *Building Research & Information*, 38(5), 458–469. Doi:10.1080/09613218.2010.494832
- van Houwelingen, Jeannet H. & van Raaij, W. F. (1989). The effect of goal-setting and daily electronic feedback on in-home energy use. *Journal of Consumer Research*, 16(1), 98–105. Doi:10.1086/209197
- Verplanken, B. & Weenig, M. W. H. (1993). Graphical energy labels and consumers' decisions about home appliances: A process tracing approach. *Journal of Economic Psychology*, 14(4), 739–752. Doi:10.1016/0167-4870(93)90019-H
- Waechter, S., Sütterlin, B. & Siegrist, M. (2015). The misleading effect of energy efficiency information on perceived energy friendliness of electric goods. *Carbon Emissions Reduction: Policies, Technologies, Monitoring, Assessment and Modeling*, 93, 193–202. Doi:10.1016/j.jclepro.2015.01.011
- Wilhite, H. & Ling, R. (1995). Measured energy savings from a more informative energy bill. *Energy and Buildings*, 22(2), 145–155. Doi:10.1016/0378-7788(94)00912-4
- Winett, R. A., Hatcher, J. W. & Fort, T. R. (1982). The effects of videotape modeling and daily feedback on residential electricity conservation, home temperature and humidity, perceived comfort, and clothing worn: Winter and summer. *Journal of Applied Behavior Analysis*, 15(3), 381–402. Doi:10.1901/jaba.1982.15-381
- Winett, R. A., Neale, M. S. & Grier, H. C. (1979). Effects of self-monitoring and feedback on residential electricity consumption. *Journal of Applied Behavior Analysis*, 12(2), 173–184. Doi:10.1901/jaba.1979.12-173

A Anhang

A.1 Schlagwortkombinationen für die Datenbanksuche

Die Datenbanken EconLit, JSTOR, und ScienceDirect, die weiter unten etwas ausführlicher beschrieben werden, wurden anhand vorab festgelegter, identischer Schlagwort-Kombinationen durchsucht, die in Tabelle A.1 aufgelistet sind. Berücksichtigt wurden dabei folgende Suchfelder:

- EconLit: Suchfelder „Title“, „Abstract“ und „Subjects (SU)“; Kategorien „Journals“ und „Working Paper“.
- JSTOR: Journal-Artikel; Fachbereiche „Business“, „Economics“, „Environmental Science“, „Environmental Studies“, „Finance“, „General Science“, „Management & Organizational Behavior“, „Psychology“ und „Sociology“.
- ScienceDirect: Suchfeld „Abstract, Title, Keywords“; ausschließlich Journal-Artikel, da keine entsprechende Kategorie für Working Paper existiert; Fachbereiche „Economics, Econometrics, Finance“, „Psychology“, „Social Sciences“, „Environment“ und „Energy“.

Tabelle A.1

Schlagwortkombinationen für die Datenbanksuche

Intervention	Schlagwortkombinationen
Sozialer Vergleich	“social norms” OR “social learning” OR “social modeling” OR “social influence” OR “peer comparison” OR “peer information” OR “comparative energy information”
Selbstbindung	“commitment” OR “pre-commitment” OR “precommitment” OR “pledge” OR “behavioral contract” OR “commitment contract” OR “commitment devices” OR “commitment approach” OR “personal commitment” OR “public commitment” OR “self-control” OR “self-regulation” AND NOT „unit commitment“
Zielsetzung	“goal setting” OR “goal-setting” OR “consumption target*” OR “consumption goal*” OR “exogenous target*” OR “exogenous goal*” OR “target request*” OR “conservation request*” OR “conservation goal*” OR “conservation target*” OR “reduction goal*” OR “reduction target*” OR “savings goal*” OR “savings target*” OR “reference goal*” OR “goal assignment” OR “self-set goal*” OR “assigned goal*” OR “reference point*”
Labeling	“energy labeling” OR “energy labelling” OR “information label*” OR “energy information” OR “energy label” OR “information acquisition” OR “information disclosure” OR “environmental certification”
Feedback	„feedback“
	AND “energy conservation” OR “energy consumption” OR “energy use” OR “energy usage” OR “energy demand*” OR “energy saving*” OR “electricity conservation” OR “electricity consumption” OR “electricity use” OR “electricity usage” OR “electricity demand*” OR “electricity saving*” OR “gas conservation” OR “gas consumption” OR “gas use” OR “gas usage” OR “gas demand*” OR “gas saving*” OR “water conservation” OR “water consumption” OR “water use” OR “water usage” OR “water demand*” OR “water saving*” OR “conservation behavior”

Die in der Tabelle genannten Suchformeln müssen für die Expertensuche bei ScienceDirect um ein tak() AND tak() AND NOT tak() ergänzt werden. Bei EconLit muss anstatt von “tak()” die Suchbegriffskette dreimal angegeben werden für abstract („AB()“), title („OR TI()“) und subject („OR SU()“). Bei JSTOR muss einzeln in „Item Title“ und „Abstract“ gesucht werden. Die angegebenen Sternchen sind nur bei EconLit und JSTOR vonnöten. Alle Datenbanken suchen automatisch nach britischer und amerikanischer Schreibweise, so zum Beispiel bei „behavior“.

EconLit ist eine von der American Economic Association erstellte Fachdatenbank für alle Fachgebiete der Wirtschaftswissenschaften und angrenzende Disziplinen. Abgebildet werden über 785 000 Titel, u.a. Artikel, Aufsätze, Bücher, Dissertationen und Arbeitspapiere. Die Datenbank wertet mehr als 750 internationale wirtschaftswissenschaftliche Zeitschriften ab 1969 aus. Ihr Schwerpunkt liegt auf US-amerikanischer Fachliteratur, doch auch deutsche und andere europäische Literaturhinweise werden aufgenommen.

JSTOR ist eine gemeinnützige Organisation, die eine Datenbank mit Artikeln aus über 2 400 Zeitschriften und über 50 000 Büchern pflegt.

ScienceDirect ist eine vom Elsevier-Verlag betriebene Online-Datenbank, die Artikel aus über 3 500 wissenschaftlichen Zeitschriften und 34 000 Büchern enthält. Insgesamt umfasst die Datenbank über 13 Millionen Artikel, die bis zum Jahr 1823 zurückreichen.

A.2 Liste der extrahierten Daten

In der folgenden Tabelle sind alle relevanten Variablen der Datenextraktion aufgelistet. In dem Datensatz entspricht eine Beobachtung einer Schätzung der jeweiligen Studie. Informationen auf Studienebene (beginnend mit s) werden jeweils nur für die erste Schätzung eines Artikels (bzw. einer Studie, falls mehrere Studien pro Artikel) festgehalten, Informationen auf Schätzungsebene (beginnend mit e) werden für jede Schätzung aufgenommen.

Tabelle A.2
Variablenliste

Variablenname	Beschreibung
Kodierungsinformationen	
s1	Identifikationsnummer in der Citavi-Datenbank
s1a	Artikelidentifikationsnummer
e2s	Fortlaufende Nummerierung der Schätzung
s3	Studien-Nummer, falls mehrere eigenständige Studien innerhalb eines Artikels
s3b	Unterschiede zwischen Studien
s5	Name des Codierers
Allgemeine Studieninformationen	
s11	Titel der Studie
s12_x	Autor/in x (Name, Vorname)
s13	Veröffentlichungsjahr
s17	Art der Veröffentlichung
s17b	Name der Zeitschrift
s18	Notizen
Studiendesign	
s21	Interventionsdesign
s21p	Seitenzahl (Interventionsdesign)
s22	Alle in der Studie betrachteten Interventionstypen
s22_x	Interventionstyp von Treatment x
s22_xo	Andere Interventionstypen von Treatment x
s23_x	Beschreibung des Treatments x
s24_x	Zeitraum des Treatments x
s22_c	Kontrollgruppentyp
s23_c	Kontrollgruppenerläuterung
s24_c	Zeitraum der Kontrollgruppenintervention
s31	Länderidentifikationscode
s39	Kommentar zu Dauer/Zeitraum
s41	Studie ist ein Experiment
s41b	Anmerkungen zum Experiment

Variablenname	Beschreibung
e42	Methode zur kausalen Analyse
e42o	Andere Methode der kausalen Analyse
e43	Anmerkungen zu besonderen Untergruppen/ Sensitivitätsanalysen
e44	Zusätzliche nicht extrahierte Untergruppe/ Sensitivitätsanalyse
Datensatz	
s61	Beobachtungseinheit (nach Studie)
e61	Beobachtungseinheit (nach Schätzung – falls unterschiedliche Beobachtungseinheiten innerhalb einer Studie)
s63	Datenerheber
s63b	Datenquelle
s64a	Erhebungskontext
s64b	Datengenerierungsprozess
s65	Kommentar zur Repräsentativität
s66_3	Gesamtzahl aller Beobachtungen in einer Studie
s66_4	Gesamtzahl aller unabhängigen Beobachtungen in einer Studie
e66_3	Gesamtzahl der Beobachtungen einer Schätzung
e66_4	Gesamtzahl der unabhängigen Beobachtungen in einer Schätzung
e67_x	Anzahl der unabhängigen Beobachtungen in Treatment x
e67_c	Anzahl der unabhängigen Beobachtungen der Kontrollgruppe
e68	Quelle der Beobachtungszahlen
e69	Anzahl der Erhebungswellen
e69b	Anmerkungen zur Anzahl der Erhebungswellen
s69_2	Kommentar zu den Daten
Abhängige Variable	
e71_1	Abhängige Variable (Text)
e71_2	Abhängige Variablen (Typ)
e71_3	Abhängige Variable (Skala)
e71_4	Abhängige Variable (Maßeinheit)
e71_5	Objektive Bewertbarkeit der abhängigen Variable
e71_6	Abhängigen Variable ist eine behauptete oder realisierte Größe
e72a	Mittelwert der abhängigen Variable (Typ)
e72b	Mittelwert der abhängigen Variable
e73a	Standardabweichung der abhängigen Variable (Typ)
e73b	Standardabweichung der abhängigen Variable
e73c	Mittelwert und Standardabweichung (Skala)
e79	Kommentar zur abhängigen Variable
Effektschätzungen	
e80_x	Art des Koeffizienten x

Variablenname	Beschreibung
e81_x	Vergleichsgruppe des Koeffizienten x (falls nicht Kontrollgruppe)
e83_xa	x.te Variable von Interesse (Text)
e83_xa_t	Zeitinteraktion von x (Text)
e83_xc	Effektgröße von x (Koeffizient)
e83_xd	Standardfehler der Effektgröße von x
e83_xe	Wert der t-Statistik der Effektgröße von x
e83_xg	p-Wert der Effektgröße von x
e83_xf	Signifikanzniveau der Effektgröße von x
e84_xa	x.te Kontrollvariable
e85	Seitenzahl (Effekte und Statistiken)
e86	Kommentar (Effekte)
e87	Bevorzugte Spezifikation (Code)
e87b	Bevorzugte Spezifikation (Erläuterung)
<i>Einschränkungen und Mechanismen</i>	
s113	Limitationen
s113b	Seitenzahl (Limitationen)
s112	Diskussion von Mechanismen
s112b	Seitenzahl (Diskussion von Mechanismen)
<i>Zusammenfassung</i>	
s111	Hauptaussage des Artikels
s111b	Seitenzahl (Hauptaussage des Artikels)
s115	Kosten-Nutzen-Analyse
s115b	Seitenzahl (Kosten der Intervention/en)
s116	Ethische Nudging-Bedenken

A.3 Hierarchisierung von Schätzern und Standardabweichungen

Tabelle A.3

Hierarchisierung von Schätzern, die zu gepoolten Effektschätzern beitragen

gesamte Studienpopulation	>	Untergruppe der Studienpopulation
gesamter Energieverbrauch	>	Energieverbrauch für bestimmte Zwecke (z.B. Heizen)
individualspezifischere fixe Effekte	>	weniger individualspezifische fixe Effekte
erweitertes Set an Kontrollvariablen	>	begrenztes Set an Kontrollvariablen
keine Interaktionsterme mit Intervention	>	Interaktionsterme mit Intervention
gesamter Beobachtungszeitraum	>	bestimmter Teil des Beobachtungszeitraums
Effekte auf Haushaltsebene	>	Effekte auf Individualebene
längerfristige Effekte	>	kurzfristige Effekte

Das Zeichen „>“ drückt die Präferenz des linken gegenüber dem rechten Fall aus. Bzgl. des letzten Punktes wurden Kurzfristeffekte nicht gänzlich verworfen, sondern für eine spätere Heterogenitätsanalyse herangezogen, im Rahmen derer versucht wird, in den Resultaten Muster entlang bestimmter Charakteristika zu finden. Gleiches wäre prinzipiell auch für die Individualebene denkbar, jedoch ist die Anzahl der Studien, die beide Ebenen betrachten, so gering, dass sich eine solche Zusatzanalyse nicht anbietet.

A.4 Berechnungsformeln der Effektgrößenstatistiken

Bei der Berechnung der beiden Effektgrößenstatistiken *SMD* und *lnRR* wird zwischen den Fällen unterschieden, in denen nur ein Schätzer pro Interventions-Ergebnisvariablen-Kombination vorliegt, und denen, wo aufgrund einer fehlenden klaren Präferenz mehrere Effektschätzungen für eine Kombination infrage kommen. Bei letzterem werden sogenannte „synthetische“ Effektgrößen berechnet, deren Formeln auf Borenstein et al. (2009b) basieren.

Tabelle A.4

Verwendete standardisierte Effektgrößenstatistiken

	Berechnung basierend auf einem einzelnen Schätzer	Synthetische Effektgrößenberechnung basierend auf mehreren Einzelschätzern
SMD		
Punktschätzer der Effektgröße	$SMD_i = \bar{X}_{i,T} - \bar{X}_{i,C} / s_i = \beta_i / s_i$	$SMD_i = \frac{1}{m} \times (\sum_{j=1}^m SMD_j)$
Standardfehler der Effektgröße	$SE(SMD)_i = SMD_i / t_i$	$SE(SMD)_i = \sqrt{\left(\frac{1}{m}\right)^2 \times (\sum_{j=1}^m SE(SMD)_j^2 + \sum_{j \neq k} (\rho_{jk} \times SE(SMD)_j \times SE(SMD)_k))}$
lnRR		
Punktschätzer der Effektgröße	$lnRR = \ln(\bar{X}_{i,T} / \bar{X}_{i,C})$	$lnRR_i = \ln\left(\frac{1}{m} \times (\sum_{j=1}^m \bar{X}_{j,t} / \bar{X}_{j,c})\right)$
Standardfehler der Effektgröße	$SE(lnRR) = \ln(\bar{X}_{i,T} / \bar{X}_{i,C}) / t$	$SE(lnRR)_i = \sqrt{\left(\frac{1}{m}\right)^2 \times (\sum_{j=1}^m SE(lnRR)_j^2 + \sum_{j \neq k} (\rho_{jk} \times SE(lnRR)_j \times SE(lnRR)_k))}$

Legende

β = Koeffizient aus Regressionsschätzungen zur Berechnung des Interventionseffekts	SE = Standardfehler (<i>standard error</i>)
C = Kontrollgruppenteilnehmer (<i>control</i>)	SMD = standardisierte Mittelwertsdifferenz (<i>standardised mean difference</i>)
i, j, k = extrahierte Schätzung	t = Student-t-Statistik des Effektschätzers
m = Anzahl der in einen synthetisierten Schätzer einfließenden Einzelschätzer	T = Treatmentgruppenteilnehmer
s = Standardabweichung der Ergebnisvariable nach der Intervention	X = Ergebnisvariable
	ρ = Korrelationskoeffizient zwischen Effektgrößenschätzungen

Unter den für die Berechnungen benötigten Variablen sind vor allem drei, die in Primärstudien nicht konsequent berichtet werden: *s*, *t*, und ρ . Da einige der Studien bereits Jahrzehnte zurückliegen, kam die Kontaktierung der Autoren nicht infrage. Stattdessen wurde (i) die Standardabweichung der Ergebnisvariable nach der Intervention *s*, falls möglich, anhand anderer Standardabweichungen approximiert (s. Tabelle unten).

Wenn (ii) die Student-t-Statistik des Effektschätzers (oder zu ihrer Berechnung verwertbare *p*-Werte) nicht aus der Studie extrahiert werden konnten, haben wir den *t*-Wert entsprechend des berichteten Signifikanzniveaus (i.d.R. durch Sternchen dargestellt) verwendet, da dies die genaueste Näherung ist, die in diesem Fall möglich ist. Dies ist ein verbreiteter Ansatz (s. z.B. Vooren et al. 2017, Karlin et al. 2015, und Rosenthal et al. 2001), der jedoch zu einem Verlust an Informationen und tendenziell zu einer Unterschätzung der Größe des Effekts führt, insbesondere wenn darüber hinaus statistisch insignifikante Ergebnisse mit $t=0$ kodiert werden wie in den zitierten Arbeiten (Pigott 2009). Stattdessen haben wir zumindest für *lnRR* einen weniger verzerrenden Ansatz gewählt: da, anders als bei *SMD*, auch die Standardabweichung selbst normiert

ist, können insignifikante Studien, die einen t -Wert berichten, herangezogen werden, um den Maximalwert von $SE(\ln RR)$ innerhalb dieser Studien zu ermitteln, $SE(\ln RR)_{max}$. Zensiert man all bei all denjenigen insignifikanten Studien, bei denen der t -Wert fehlt, den Standardfehler auf diesen Wert ($SE(\ln RR)_{insignifikant_t\text{-fehlend}} \triangleq SE(\ln RR)_{max}$) und formt die Gleichung aus der Tabelle um, kann man ebenfalls einen weiterhin konservativen, aber vermutlich weniger stark verzerrten t -Wert ermitteln.

In den wenigen Fällen, in denen synthetische Effektgrößen berechnet werden mussten, ist zudem an sich (iii) der Korrelationskoeffizient zwischen Effektgrößenschätzungen, ρ , vonnöten. Dieser Koeffizient ist jedoch unbekannt und für die einbezogenen Studien kaum plausibel approximierbar. Es wird daher von einem Koeffizienten von 1 ausgegangen, d.h. perfekter Korrelation. Diese Annahme ist ebenfalls konservativ, da dadurch die Varianz vermutlich überschätzt und somit die Genauigkeit des Schätzers unterschätzt wird.

Tabelle A.5

Hierarchisierung von Standardabweichungen der Ergebnisvariable zur Berechnung der standardisierten Effektgröße

Hierarchie	Symbol	Typ der Standardabweichung (sd)	Referenz
	$s_{i,p,e}$	gepoolte sd (post)	Hedges (1981)
=	$s_i(\varepsilon)$	sd des Fehlerterms	Keef und Roberts (2004)
>	$s_{i,c,e}$	sd der Kontrollgruppe (post)	Glas (1976)
>	$s_{i,e}$	Gesamt-Standardabweichung (post)	-
>	s_i	Gesamt-Standardabweichung (cross-panel)	-
>	$s_{i,b}$	Gesamt-Standardabweichung (vorher)	-
>	$s_{i,c,b}$	sd der Kontrollgruppe (vorher)	-
>	$s_{i,t,b}$	sd der Treatmentgruppe (vorher)	-
>	$s_{i,\checkmark}$	Borenstein-Angleichung der gepoolten sd	basierend auf Borenstein et al. (2009b: 22)
>	s_i'	Gesamt-Standardabweichung auf der log-Skala transformiert (cross-panel)	auf der Grundlage von Higgins et al. (2008: 6075)

Das Zeichen „>“ oder „=“ drückt die Präferenz (bzw. Gleichbewertung) der oberen Standardabweichung zu der in der jeweiligen Zeile aufgeführten Standardabweichung aus.

A.5 Anleitung zum Lesen und Interpretieren der *Forest Plots*

In einem *Forest Plot* wie dem in Abbildung 4 sind alle jeweils aufgenommenen Studien untereinander aufgelistet. Da mitunter mehrere unabhängige Effektschätzer aus demselben Artikel auftauchen, werden diese wie folgt gekennzeichnet:

- stX für Studie X innerhalb eines Artikels
- GX für Treatmentgruppe X , falls also mehrere separate Gruppen unterschiedlichen Interventionen (oder Interventionskombinationen) unterzogen wurden
- $c(X)$ für den Fall, dass nicht mit einer reinen Kontrollgruppe verglichen wird, sondern mit einem anderen Treatment
- *aggr E.*, falls die Studie Strom- und Gasverbrauch für die entsprechende Intervention nur kombiniert betrachtet.

Darüber hinaus sind die Effekte nochmals aufgeteilt in solche, bei denen Studienteilnehmer nur die jeweilige Intervention durchlaufen („reiner Interventionseffekt“) und solchen, bei denen die jeweils untersuchte Intervention noch mit anderen kombiniert wurde („Effekt aus Kombination mit anderen Interventionen“, s.a. Abbildung 3 in Kapitel 4.3). Neben jeder Studie wird rechts eine horizontale Linie und ein Quadrat dargestellt. Die Mitte des Quadrats zeigt den punktgenauen Wert des Ergebnisses der jeweiligen Studie an. Ein $InRR$ -Wert von 1 bedeutet, dass der Verbrauch zwischen Treatment- und Kontrollgruppe gleich ist, somit als nicht von der Intervention beeinflusst angesehen werden kann. Dies wird durch die vertikale Mittellinie als „Kein-Effekt-Linie“ dargestellt. Abweichungen von 1 sind prozentual zu verstehen, ein Wert von 0.95 entspricht somit einer Verbrauchsreduzierung um 5 Prozent. Je weiter ein Quadrat von der Mittellinie entfernt ist, desto stärker war der Effekt in der Studie.

Die Größe des Quadrats repräsentiert die Bedeutung (das „Gewicht“) der Studie für das Gesamtergebnis. Durch die unterschiedliche Gewichtung ist es möglich, in einer Meta-Analyse Studien unterschiedlicher Größe und Präzision zu berücksichtigen. Technisch entspricht das Gewicht der inversen Varianz des Effektes, d. h. je präziser in einer Studie der Effekt bestimmt werden kann, desto mehr Gewicht wird auf diese Studie in der Meta-Analyse gelegt.

Die Länge der horizontalen Linien links und rechts der Quadrate zeigt das Konfidenzintervall an. Hier kommt wieder die „Kein-Effekt-Linie“ ins Spiel. Wenn die Linie der Einzelstudie diese Mittellinie schneidet, ist das Ergebnis der Einzelstudie nicht signifikant (5 %-Niveau). Befindet sich dagegen die komplette Linie auf einer Seite der Mittellinie, handelt es sich um einen signifikanten Effekt. Konfidenzintervalle werden in der rechten Spalte in Klammern hinter dem Punktschätzer des Effektes angegeben.

Die Raute unter den Einzelstudien zeigt das Gesamtergebnis der gepoolten Effekte in der Meta-Analyse. Durch die Lage des Mittelpunktes ist sofort ersichtlich, ob die Intervention tatsächlich zu einer Verbrauchsreduzierung geführt hat. Das Konfidenzintervall des Gesamtergebnisses wird durch die äußeren Punkte der Raute dargestellt. Durch die gemeinsame Betrachtung aller relevanten Studien in einer Meta-Analyse wird die Stichprobe größer als in den Einzelstudien, wodurch die Präzision der Abschätzung des Gesamteffektes erhöht wird, weswegen die Raute in der Regel relativ schmal ist. Ein eigenes statistisches Signifikanzniveau in Form eines p -Wertes wird zudem auf der linken Seite aufgezeigt. Ein p -Wert kleiner als 10 bzw. 5 Prozent wird üblicherweise als statistisch signifikant interpretiert, es ist somit äußerst unwahrscheinlich, dass der beobachtete Unterschied dem Zufall zuzuschreiben ist. Neben diesem Wert wird abschließend noch ein sogenannter I^2 -Wert („I-squared“) angegeben. Er vermittelt dem Leser, wie ähnlich (homogen) die zugrundeliegenden Studien selber sind, inwiefern sie also tatsächlich „in einen Topf

geworfen“ werden könne („Äpfel und Birnen“-Problem). Bei einem I^2 -Wert von 0 % wäre die Variabilität der Studienergebnisse vollständig auf Zufallsschwankungen zurückzuführen. Bei einem Wert von 100 % wäre die Variabilität allein durch Unterschiede zwischen den Studien zu erklären. Große Unterschiede in den Effektgrößen oder Konfidenzintervallen deuten auf systematische Unterschiede zwischen den Studien hin. Das schränkt die Vertrauenswürdigkeit des gepoolten Gesamtergebnisses ein (Higgins et al. 2003). Higgins und Green (2011) schlagen vor, einen Wert über 50 Prozent als Indiz für substantielle Heterogenität zu werten.¹³

¹³ Die Erläuterungen lehnen sich an <http://www.dbm-agentur.eu/wie-liest-man-einen-forest-plot.html> und Weckmann et al. (2015) an.



Das RWI wird vom Bund und vom Land
Nordrhein-Westfalen gefördert.