

Nudges, Transparenz und Autonomie – Eine normativ gehaltvolle Kategorisierung von Maßnahmen des Nudging

LUDGER HEIDBRINK UND ANDREA KLONSCHINSKI

Ludger Heidbrink, Christian-Albrechts-Universität zu Kiel, E-Mail: heidbrink@philsem.uni-kiel.de
Andrea Klonschinski, Christian-Albrechts-Universität zu Kiel, E-Mail: klonschinski@philsem.uni-kiel.de

Zusammenfassung: Die Debatte um Nudging wird durch die große Bandbreite dessen, was als „Nudge“ betrachtet wird, erschwert. Die Einschränkung des Begriffs auf Maßnahmen, die Rationalitätsdefizite des Individuums ausnutzen, hat sich in der Debatte nicht durchgesetzt. Da die so definierten Nudges zudem als ethisch problematisch zu beurteilen sind, ist eine differenzierte Betrachtung des Phänomens angezeigt. In dem Beitrag werden Nudges anhand ihrer Transparenz einerseits sowie der Frage, ob sie nur auf das automatische oder auch auf das reflexive Denksystem wirken andererseits, kategorisiert. Argumentiert wird in diesem Beitrag, dass Nudges unproblematisch sind, wenn sie eigenverantwortliches Handeln zulassen und auf ein vom Individuum selbst als solches identifiziertes Entscheidungsproblem reagieren. Dies ist der Fall, wenn Nudges (i) transparent auf das reflexive System als Mittel gegen Unachtsamkeit oder Vergesslichkeit wirken (zum Beispiel Warnhinweise), (ii) eine edukative Funktion haben (zum Beispiel Art der Angabe von Wahrscheinlichkeiten) oder (iii) auf Probleme mangelnder Selbstkontrolle reagieren, die vom Individuum als nicht in seine übergeordnete Präferenzstruktur passend bewertet werden (zum Beispiel Gesundheitsvorsorge).

Summary: The debate on the legitimacy of nudging is hampered by the absence of a clear definition of what a nudge is. Given that the attempt to limit the term “nudge” to interventions exploiting cognitive biases has not been accepted in the literature and since the respective interventions lack legitimacy, a more fine-grained conceptualization of nudges seems warranted. The paper offers a categorization of nudges on the basis of their transparency on and the question of whether the intervention aims at the automatic system only or tackles the reflexive system as well. We argue that nudges are legitimate when they allow for autonomous choice and aim at a decision problem which is actually identified as such by the individual herself. This is the case when (i) nudges are transparent, tackle the reflective system, and help the individual dealing with issues of attention and obliviousness (for example warnings), (ii) have an educative function, and (iii) when they respond to issues of self-control, which the individual regards as not fitting into her long-term preferences and her overarching scheme of life.

- JEL classification: D11, E21, I31, Y80, Z18
- Keywords: Libertarian paternalism, nudge, two systems approach, autonomy, self-control

I Einleitung

Der libertäre Paternalismus, wie er insbesondere von Richard H. Thaler und Cass R. Sunstein vertreten wird, beruht auf der alltäglichen und durch verhaltensökonomische Studien belegten Beobachtung, dass Individuen in ihrer Wahrnehmung und Entscheidungsfindung kognitiven Verzerrungen unterliegen, häufig nicht entsprechend ihrer langfristigen, konsistenten und voll informierten Präferenzen handeln, willensschwach und vergesslich sind. Durch eine entsprechend gewählte Präsentation der Entscheidungsoptionen, so die Grundidee des libertären Paternalismus, könne den Individuen jedoch dazu verholfen werden, ihren eigentlichen, reflektierten und rationalen Präferenzen gemäß zu handeln. Derartige Maßnahmen des Nudgings seien paternalistisch, insofern sie auf das Wohl der Betroffenen abstellen, wie diese es definieren, aber zugleich libertär, da sie den Individuen keiner Entscheidungsoptionen berauben (Sunstein und Thaler 2003, Thaler und Sunstein 2003, 2011). Während Nudges theoretisch zu allen möglichen Zwecken eingesetzt werden können, die nicht dem Wohl des Betroffenen (oder dem Wohl Dritter) dienen, betrachten wir im Folgenden ausschließlich Nudges im Rahmen des libertären Paternalismus, der dem Slogan „*nudge for good*“ (Thaler 2015: 345) verschrieben ist.

Während sich der libertäre Paternalismus einerseits in Politik und Wissenschaft einer großen Beliebtheit erfreut, ist er andererseits Gegenstand anhaltender Kritik. Neben den Fragen, ob Individuen überhaupt über so etwas wie eine wahre, konsistente Präferenzordnung verfügen, und wie diese theoretisch und praktisch zu eruieren sei (Sugden 2008, Grüne-Yanoff 2012, McQuillin und Sugden 2012, Infante et al. 2016), ist es von einem normativ-ethischen Standpunkt aus insbesondere als problematisch zu beurteilen, dass gezielt auf unbewusste Denkprozesse abstellende Nudges die Autonomie des Individuums untergraben können (Hausman und Welch 2010, Rebo-

Kasten

Sunsteins Beispiele für Nudges

Sunstein (2016a) nennt beispielsweise folgende Nudges:

„*a failure to warn is a failure to nudge*“, „*[d]isclosure of relevant information is a nudge*“, „*[a] recommendation is a nudge*“, „*[u]ses of color and large fonts can be nudges*“, „*educational campaigns [nudge]*“ und „*salience is a nudge*“; hinsichtlich der Akteure schreibt er: „*Hitler nudged; so did Stalin. Terrorists nudge*“, „*[n]ature nudges*“ ebenso wie „*the common law of contract, tort, and property*“ (Sunstein 2016a: 16–35).

nato 2012, White 2013, Wilkinson 2013, Baldwin 2014). Trotz dieser Kritik scheint es andererseits nicht zielführend, das Kind mit dem Bade auszuschütten und Nudging generell als illegitim zu verurteilen; so lassen sich durchaus Argumente dafür anführen, dass einige dieser Maßnahmen aus autonomieorientierter Perspektive unproblematisch sind und den Individuen klare Vorteile bringen.

Eine Beurteilung der ethischen Legitimierbarkeit des libertären Paternalismus wird dabei allerdings durch die große Bandbreite an vorgeschlagenen Nudges erschwert; geht in der Debatte

doch nahezu jede Art der nicht mit Zwang verbundenen Verhaltensbeeinflussung – sei es durch Information, monetäre Anreize, das Wetter oder den eigenen Wecker – als Nudge durch (siehe etwa Thaler und Sunstein 2011: 297ff., Sunstein 2013: 38f., 2016a: 26).

Auch sind diese Verhaltensbeeinflussungen, anders als es der Begriff des libertären Paternalismus ankündigt, keineswegs notwendigerweise paternalistisch, sondern umfassen etwa auch Maßnahmen, die auf Umweltschutz (Thaler und Sunstein 2011: 251 ff.) oder das Wohl anderer abstellen. Ein Beispiel für Letzteres sind etwa die Überlegungen zur Erhöhung der Organspendebereitschaft in der Gesellschaft (Thaler und Sunstein 2011: 240 ff., Sunstein 2013: 103). Ange-sichts dieser Vielfalt der Maßnahmen gilt es im Sinne einer differenzierten normativ-ethischen Bewertung verschiedene Arten des Nudgings zu kategorisieren und Nudges damit zu kontextualisieren.¹ Dies leistet der vorliegende Beitrag. Es sei angemerkt, dass wir uns im Folgenden auf absichtlich gesetzte Nudges konzentrieren und etwa Verhaltensbeeinflussungen durch das Wetter unberücksichtigt lassen, da die Diskussion der ethischen Legitimierbarkeit von Nudging von handelnden Subjekten ausgehen muss. Sunstein (2016a: 36) selbst spricht sich dezidiert gegen diese Einschränkung aus.

Dabei gehen wir vor wie folgt. Zunächst betrachten wir den prominenten Versuch, Nudges durch ihre Wirkung auf das sogenannte automatische System des Denkens trennscharf von anderen regulatorischen Maßnahmen abzugrenzen (Abschnitt 2). Nudges, so die Idee, können ihre Wirkung nur deswegen entfalten, weil Menschen kognitiven Verzerrungen unterliegen; auf vollständig rationale Wesen hätten sie hingegen keinen Einfluss. Werden Nudges auf diese Art und Weise als Rationalitätsdefizite ausnutzende und im Ergebnis behebende Maßnahmen verstanden, so ist ihre Legitimationsbasis indes schwach. Da sich diese Einengung des Begriffs zudem nicht hat durchsetzen können, ist es sinnvoll, die große Gruppe möglicher Nudges anders zu kategorisieren. Die unserer Ansicht nach geeignetste und auch normativ-gehaltvollste Kategorisierung stammt dabei von Hansen und Jespersen (2013), die Nudges anhand des Kriteriums der Transparenz der Maßnahme einerseits und der Frage, ob die Maßnahme allein auf das automatische oder auch auf das reflexive Denksystem abstellt andererseits, differenzieren. Der vorliegende Beitrag erläutert und ergänzt dieses Schema, indem er drei zentrale Gruppen vermeintlicher Entscheidungsfehler identifiziert und diese in die Kategorisierung von Hansen und Jespersen einordnet (Abschnitt 3). Die vorgeschlagene Unterscheidung erlaubt eine differenzierte Beantwortung der Frage, welche Art von Nudges als ethisch legitim bezeichnet werden kann und welche nicht. Der Beitrag nimmt damit insgesamt eine wichtige begriffliche und normative Differenzierung vor und bietet einen Rahmen für weitere Differenzierungen und Diskussionen insbesondere von Nudging in der politischen Planung, Lenkung und Gestaltung (Abschnitt 4).

1 Dies fordert mittlerweile auch Sunstein (2016a: 26, 2016b: 21). Zur Kontextualisierung und Differenzierung von Nudges existieren in der Literatur Ansätze mit unterschiedlicher Schwerpunktsetzung (Hausman und Welch 2010, Rebonato 2012, Hansen und Jespersen 2013, Baldwin 2014, Hanna 2015, Hansen 2016, van Aaken 2016).

2 Definition, Kategorisierung und Legitimität von Nudges

2.1 Die verhaltensökonomische Definition und ihre normative Basis

Wie also lässt sich Nudging von anderen Verhaltensbeeinflussungen abgrenzen? Thaler und Sunstein verstehen unter Nudge „alle Maßnahmen, mit denen Entscheidungsarchitekten das Verhalten von Menschen in vorhersagbarer Weise verändern können, ohne irgendwelche Optionen auszuschließen oder wirtschaftliche Anreize stark zu verändern. Ein Nudge muss zugleich leicht und ohne großen Aufwand zu umgehen sein“ (Thaler und Sunstein 2011: 15). Diese negative Definition dessen, was ein Nudge nicht ist, wird durch den Rekurs auf eine bestimmte Wirkungsweise von Nudging positiv spezifiziert; demnach sei ein Nudge „jeder Faktor, der das Verhalten von Humans signifikant verändert, während er von Econ ignoriert würde“ (Thaler und Sunstein 2011: 19). Ein sogenannter „Econ“, der durch vollständige Informiertheit und Rationalität gekennzeichnet ist, dürfte sich von Nudges also nicht beeinflussen lassen, wohingegen diese Maßnahmen durchaus Einfluss auf den von kognitiven Unzulänglichkeiten geplagten „Human“ haben. Diese Differenzierung beruht auf der Unterscheidung von zwei Denksystemen, dem automatischen System einerseits und dem reflexiven andererseits, die es im Folgenden zu erläutern gilt (Thaler und Sunstein 2011: 33ff., Kahneman 2011).

Während das reflexive System für das bewusste, konzentrierte und langsame Nachdenken zuständig ist, arbeitet das automatische System unbewusst, mühelos und schnell (kritisch dazu etwa Rebonato 2012: 35, Gigerenzer 2015, Infante et al. 2016). Automatisches Denken ist für den Alltag zwar unerlässlich, aber gerade aufgrund seines weitgehend unkontrollierten Funktionierens anfällig für Fehler, wie die Verhaltensökonomik ausführlich demonstriert. Da reale Menschen Humans und keine Econ seien, so Thaler (2015: 325), machen sie systematische Fehler. Nudges zielen auf die Korrektur dieser Fehler und schubsen die Individuen zu den Entscheidungen, die sie selbst treffen würden „wenn sie richtig aufgepasst hätten, umfassend informiert wären und unbegrenzte kognitive Fähigkeiten sowie absolute Selbstkontrolle besäßen“ (Thaler und Sunstein 2011: 15).

Insofern mit Econ auf das in wirtschaftswissenschaftlichen Lehrbüchern dargestellte rationale Wesen der Ökonomik verwiesen wird sowie aufgrund anderer Formulierungen von Thaler und Sunstein (2003: 176, Sunstein und Thaler 2003: 1168, insbesondere Thaler 2015: 29, 57) lässt sich schließen, dass diese hier auf die moderne Rationalitätskonzeption der Ökonomik, die sogenannte Erwartungsnutzentheorie (dazu siehe Hausman 2012: 37–42) abstehen (Rebonato 2012: 36, mit Bezug zum libertären Paternalismus siehe Klonschinski und Wündisch 2016). Thaler und Sunstein gehen folglich davon aus, dass das reflexive Denken diesem Rationalitätsmodell folgt. Damit ist ein klares Abgrenzungskriterium für Nudges gefunden, welches zudem ein wahrhaft originelles Kernstück des libertären Paternalismus ausmacht, wie auch Rebonato (2012: 84) hervorhebt: Nudges nutzen gezielt kognitive Mängel der Individuen aus, um deren Entscheidungen zu verbessern (Hausman und Welch 2010: 126, Hansen 2016: 8). Sieht man von dieser verhaltensökonomisch basierten Art der Nudges ab, so würden die unter diesem Begriff firmierenden Maßnahmen der Beeinflussung in Regierung und Marketing seit Jahrzehnten verwendet und der libertäre Paternalismus könnte keinerlei Anspruch auf Originalität erheben (Rebonato 2012: 64, Gigerenzer 2015: 362).

Dieser Begriffsbestimmung zufolge würde etwa reine Information keinen Nudge darstellen, da sie auf das reflexive System wirkt und auch einen Econ beeinflusste; dasselbe gilt für Anreize. Unterschiedliche Darstellungen (*Framing*) ein und derselben Aussage hingegen – zum Beispiel

die Reihenfolge relevanter Informationen – dürfte den Econ nicht beeinflussen, wirkt sich aber auf den Human aus – ein typischer Nudge.

Ein zentrales Problem der so definierten Nudges ist allerdings die Annahme, dass das vollständig rationale Verhalten des Econ als idealer Maßstab für Verhaltensweisen angenommen wird, die das eigene Wohlergehen befördern (Berg und Gigerenzer 2010: 147, Rebonato 2012: 36, Neumann 2013: 92 ff., Gigerenzer 2015: 365, Infante et al. 2016). Diese Prämisse ist in mehrererlei Hinsicht problematisch. Erstens bleibt unklar, was mit dem Idealwert „vollständiger Information“ oder „voller Aufmerksamkeit“ genau gemeint ist. Auch ist offensichtlich, dass kein menschliches Wesen über grenzenlose kognitive Fähigkeiten und Selbstkontrolle verfügt, sodass die Rekonstruktion der *eigentlichen Präferenzen* der Individuen mit dieser Vorannahme unmöglich gemacht wird. Dies gilt nicht nur für potentielle Entscheidungsarchitekten, sondern auch für die betreffenden Individuen selbst (Sugden 2008: 232, 2009: 370, Grüne-Yanoff 2012: 642 f., McQuillin und Sugden 2012: 560 f., White 2013: 76–78).

Zweitens ist angesichts der lediglich auf konsistente Präferenzen abstellenden Rationalitätskonzeption der modernen Ökonomik keine Aussage über die Ziele der Individuen möglich. Insbesondere folgt daraus nicht, dass Individuen präferieren, was in irgendeinem substantiellen Sinne „gut“ für sie ist und ihnen etwa mehr „*health, wealth and happiness*“ verschafft, wie es der Untertitel von Thaler und Sunstein (2009) verspricht. Die Frage, ob ein Econ das Risiko eines kürzeren Lebens für ein genussvoller Leben eingeht oder ein längeres Leben mit mehr Versagungen präferiert, kann nicht ohne zusätzliche Annahmen beantwortet werden. Der Entscheidungsarchitekt kann mithin nicht wissen, *was* ein rationales Individuum will (White 2013: 61 ff., Lusk 2014: 367, ausführlicher Klonschinski und Wündisch 2016). Drittens konnte bislang auch empirisch nicht gezeigt werden, dass Individuen Nachteile in Form von Glücks- oder Einkommenseinbußen aus „irrationalem“ Verhalten erwachsen (Berg und Gigerenzer 2010: 148, Gigerenzer 2015: 370). Relevant ist hier auch der Einwand, dass die Ergebnisse der Verhaltensökonomik nicht ohne weiteres auf die Welt außerhalb des Labors übertragbar sind (Neumann 2013: 86 ff., Lusk 2014: 362 ff.).

Der allgemeine Verweis auf Rationalitätsdefizite der Individuen stellt folglich aus mehreren Gründen keine legitime Basis für Nudging dar. Weder ist empirisch klar, ob Abweichungen vom Rationalitätsideal „schlecht“ für das Individuum sind, noch kann der Entscheidungsarchitekt wissen, was die Individuen wollen beziehungsweise gewollt hätten, wenn sie rational gewesen wären. Unabhängig von diesem normativen Problem hat sich die Einengung des Begriffs „Nudging“ auf Maßnahmen, die das automatische System „ausnutzen“ in der Debatte nicht durchsetzen können (Baldwin 2014: 835). In einer jüngeren Publikation weist Sunstein die verhaltensökonomische Definition dezidiert zurück: „*But (and this is an important point) a behavioral bias is not a necessary justification for a nudge, and nudges need not target or exploit System 1 in any way*“ (Sunstein 2016b: 35 f.).

Vor dem Hintergrund dieser Überlegungen und insbesondere, um nicht in einen harten Paternalismus zu verfallen, scheint es sinnvoll, nur diejenigen Nudges als legitim zu betrachten, die nicht auf Mutmaßungen darüber, was „gut“ für das Individuum sei, angewiesen sind. An dieser Stelle lässt sich auf die Unterscheidung zwischen Präferenz- oder Zielpaternalismus einerseits und kognitivem oder Mittelpaternalismus andererseits von van Aaken (2016: 171) zurückgreifen. Während der Zielpaternalismus die Individuen zu einer substantiellen Konzeption des guten Lebens hinführen will, nimmt der Mittelpaternalismus die Präferenzen der Individuen als gegeben an und hilft ihnen, diese angesichts von kognitiven oder motivationalen Schwierigkeiten zu

realisieren (van Aaken 2016: 171). Entsprechende Nudges sind dann insbesondere diejenigen, die einen solchen Grad an Transparenz aufweisen, dass es tatsächlich leicht vermieden werden kann, ihnen Folge zu leisten. Zudem müssen sie zusätzlich das Kriterium erfüllen, einem Entscheidungs- oder Verhaltensproblem Abhilfe zu verschaffen, welches das Individuum explizit selbst als mit seinen eigenen, langfristigen Präferenzen unvereinbar bewertet.² Diese allgemeinen und spezifizierungsbedürftigen Überlegungen sollen im Folgenden auf der Basis einer Kategorisierung von Nudges, wie sie von Hansen und Jespersen (2013) vorgeschlagen worden ist, systematisiert und erläutert werden.

2.2 Transparenz und Reflexion

Hansen und Jespersens (2013) Systematisierung von Maßnahmen des Nudging basiert auf dem Kriterium der Transparenz einerseits sowie der Wirkungsweise der Maßnahme auf die beiden Denksysteme andererseits. Transparent sei ein Nudge, wenn vernünftigerweise erwartet werden kann, dass die Zielperson sowohl die verhaltensbeeinflussende Intention hinter der Maßnahme als auch die verwendeten Mittel der Beeinflussung als solche erkennt (Hansen und Jespersen 2013: 17). Als Beispiele nennen die Autoren unter anderem die Fliege im Urinal des Schiphol Airport, die zum gezielten Urinieren einlädt oder den Hinweis „Look-right“ auf Londoner Straßen, der Fußgänger auf den Linksverkehr hinweist (Hansen und Jespersen 2013: 18). Als einen intransparenten Nudge hingegen bezeichnen Hansen und Jespersen (2013: 18) eine Maßnahme, hinsichtlich derer nicht vernünftigerweise von der Zielperson erwartet werden könnte, dass diese die Intention der Beeinflussung und die dafür verwendeten Mittel als solche erkennt. Beispiele umfassen die Verkleinerung von Tellern, um Individuen zu weniger Kalorienaufnahme zu bewegen, oder das Framing alternativer medizinischer Behandlungsmethoden. Mit Blumenthal-Barby (2012: 357) ließen sich an dieser Stelle zwei weitere relevante Faktoren hinzufügen, die die Transparenz einer Maßnahme beeinflussen: die Beziehung zwischen den Akteuren und deren legitime Erwartung aneinander. Während etwa von der eigenen Mutter erwartet werden könnte, dass sie die Gesundheitsrisiken einer bestimmten Tätigkeit aus Sorge übertrieben darstellt, sei dies bei einem Arzt nicht der Fall (Blumenthal-Barby 2012: 363). Ein und dieselbe Maßnahme kann damit je nach Kontext als transparent oder als intransparent gelten.

Zusätzlich differenzieren Hansen und Jespersen (2013: 14) zwei Typen von Nudges danach, ob sie allein auf das automatische oder auch auf das reflexive Denksystem wirken:

Typ 1 Nudges zielen darauf ab, die intuitive Aufmerksamkeit und die unbewussten Prämissen des reflexiven Denkens zu beeinflussen. Sie wirken also *unmittelbar* auf das automatische System, indem sie etwa Aufmerksamkeit erregen, werden aber erst nach einem reflexiven Prozess in Handlungen umgesetzt. Dies passiert etwa bei der bereits erwähnten Fliege im Urinal.

Typ 2 Nudges hingegen stellen auf eine automatische Änderung des *Verhaltens* – im Gegensatz zum bewussten *Handeln* – ohne „Umweg“ über das reflexive System ab. Hier wären die Streifen auf dem Lake Shore Drive, um die Geschwindigkeit zu reduzieren, oder die Variation von Tellergrößen, um die Essensmenge zu beeinflussen, zu nennen. Die Betroffenen überlegen nicht erst, ob sie langsamer fahren, mehr oder weniger essen sollen; sie tun es einfach.

2 Hinzu treten Nudges, die darauf abzielen, gesellschaftlich anerkannte oder sogar rechtlich kodifizierte Ziele zu fördern, wie etwa Umweltschutz. Hier wäre allerdings genauer zu bestimmen, was „gesellschaftlich anerkannt“ bedeutet.

Übersicht 1

Systematisierung von Nudges

	Transparent	Intransparent
Typ 1: Reflexives System	Nudge I zum Beispiel: · Fliege im Urinal, · automatische Erinnerungen, · Warnsignale, · Ampeln auf Lebensmitteln, · Bilder auf Zigarettenpackungen	Nudge II zum Beispiel: · Anchoring, · Framing
	Fehlerquellen: Unachtsamkeit, Vergesslichkeit → Nudges als Hilfsmittel	Fehlerquellen: unangemessen gewählte Heuristiken → edukative Nudges
Typ 2: Automatisches System	Nudge III zum Beispiel: · Signalfarben, · Änderungen von Standardeinstellungen, die man relativ einfach bemerkt (Drucker, Sparpläne, Betriebsrenten)	Nudge IV zum Beispiel: · Tellergröße, · Anordnung von Produkten in der Cafeteria, · Vorgabe der Organspende (Widerspruchslösung)
	Fehlerquellen: mangelnde Selbstkontrolle, Willensschwäche, Trägheit → Nudges als Mittel vernünftiger Selbstbindung	

Quelle: Eigene Darstellung 2017 in Anlehnung an Hansen und Jespersen (2013: 23).

Auf Basis dieser Differenzierungen lassen sich Nudges des Typs 1 und 2 in einer Matrix anordnen (Übersicht 1). Im Folgenden erläutern wir die Nudges in den vier Feldern, ordnen ihnen jeweils eine typische „Fehlerquelle“ zu und diskutieren die normativ-ethische Vertretbarkeit entsprechender Maßnahmen.

3 Systematisierung von Nudges und Entscheidungsfehlern

3.1 Nudge I: Unachtsamkeit und Vergesslichkeit

Nudges des Typs I sind dadurch gekennzeichnet, dass sie vom Individuum relativ leicht als intendierte Maßnahmen, die ihre Handlungen beeinflussen sollen, erkannt werden und dass ein reflexiver Prozess eintreten muss, bevor der Nudge wirken kann. Wiederum lassen sich die Beispiele der Fliege im Urinal und die Abbildungen auf Zigarettenpackungen heranziehen. In beiden Fällen wird zunächst Aufmerksamkeit erregt und, im Falle der Zigaretten, eine bestimmte mögliche Konsequenz des Handelns gezeigt. Das darauf folgende Zielen auf die Fliege und der Kauf oder Nichtkauf der Zigaretten basieren hingegen auf einer bewussten Entscheidung, in der die Tatsache berücksichtigt wird oder werden kann, dass und wie jemand beeinflusst wird. Weitere Beispiele umfassen den Alarm im Auto, wenn kein Sicherheitsgurt angelegt wurde, die Aufkleber für verschiedene Wochentage auf der Antibabypille, Erinnerungsanrufe oder Kurznachrichten für den nächsten Zahnarzttermin und so weiter. Zentral ist dabei, dass sich das Individuum im

Zuge dieser Reflexion auch aktiv gegen die von Entscheidungsarchitekten intendierte Handlung entscheiden kann (Hansen und Jespersen 2013: 21).

Nudges vom Typ Nudge I sind also weder manipulativ im Sinne von unbewusst beeinflussend noch schränken sie die Autonomie des Handelnden ein. Sie können insofern als transparente und autonomiefördernde Form des Paternalismus betrachtet werden, als sie den Akteuren helfen, ihre tatsächlichen eigenen Ziele – zum Beispiel mit dem Rauchen aufzuhören oder Termine einzuhalten – zu verwirklichen. Die Nudges setzen dabei keine Konzeption dessen voraus, was gut für das Individuum sei, sondern stellen ein Instrument dar, langfristige Präferenzen und das reflexive Denken zu aktivieren, um so gegebenenfalls die eigene Unachtsamkeit und Vergesslichkeit zu überwinden (Hansen und Jespersen 2013: 19).

3.2 Nudge II: Unangemessen gewählte Heuristiken

Maßnahmen vom Typ Nudge II beinhalten zwar auch einen reflexiven Prozess; anders als beim Typ Nudge I umfasst dieser jedoch nicht das Bewusstsein über die Tatsache und Art der Beeinflussung des Verhaltens selbst. Dies ist etwa der Fall bei verschiedenen Framings von Wahrscheinlichkeiten. So wirkt es sich beispielsweise unterschiedlich auf die Entscheidung des Patienten aus, ob ein Arzt das Risiko einer Operation in Form von Überlebens- oder von Sterbensraten formuliert (Thaler und Sunstein 2011: 57 f., Kahneman 2011: 367 f.). In die Kategorie von Nudge II fallen auch Entscheidungen, die auf sogenannten Ankereffekten (*Anchoring*) basieren, bei denen bestimmte, den Individuen präsentierte Zahlen als Anker der folgenden Antwort dienen. Der Ankereffekt lässt sich besonders gut an einem Experiment von Kahneman und Tversky illustrieren (Kahneman 2011: 119). Sie baten ihre Versuchspersonen, an einem Glücksrad zu drehen und das Ergebnis aufzuschreiben. Das Rad trug die Zahlen 0 bis 100, war aber derart manipuliert, dass es auf 10 oder 65 stehen blieb. Im Anschluss stellten sie den Teilnehmern die folgenden Fragen: Ist der Anteil afrikanischer Nationen innerhalb der Vereinten Nationen größer oder kleiner als die Zahl, die Sie gerade notiert haben? Wie hoch schätzen Sie den Prozentsatz afrikanischer Nationen innerhalb der Vereinten Nationen? (Kahneman 2011: 119). Es zeigte sich, dass die Schätzung der Anzahl afrikanischer Länder sich systematisch an der zuvor notierten Zahl orientierte (diejenigen, die 10 notiert hatten, schätzten die Zahl durchschnittlich auf 25 Prozent, diejenigen mit einer 65 auf 45 Prozent), obwohl diese natürlich in keinerlei Zusammenhang mit der gestellten Frage steht. Der Einfluss derartiger Anker kann für das Nudging fruchtbar gemacht werden, indem die Aufmerksamkeit von Personen auf eine bestimmte Zahl gelenkt wird, bevor sie etwa kundtun, wieviel sie bereit sind zu zahlen oder zu spenden.

Die genannten Nudges haben gemeinsam, dass sie zwar eine bewusste, reflektierte Entscheidung verlangen, vom Individuum aber kaum als das Verhalten beeinflussende Maßnahme identifiziert werden können. Dies, so unsere Hypothese, liegt daran, dass die entsprechenden Verhaltens- oder Entscheidungsmuster aus tief verankerten und unbewussten Heuristiken resultieren. Die genannten Beispiele zeigen, dass durch die Wahl einer bestimmten Heuristik bedingte Entscheidungen – etwa über die Höhe einer Spende oder die Bereitschaft, ein bestimmtes Risiko einzugehen – häufig gar keine Unterscheidung in „richtig“ oder „falsch“ zulassen. Es ist erst einmal nur festzustellen, dass Entscheidungen abhängig von der Art der verwendeten Heuristik sind. Daher ist es nicht angemessen, die Verwendung bestimmter Heuristiken per se als Ursache von Entscheidungsfehlern (Thaler und Sunstein 2011: 31) zu betrachten und daraus die Legitimität von Nudging abzuleiten (Rachlinski 2003).

Wenn aus unangemessen gewählten Heuristiken dennoch Entscheidungsfehler, beispielsweise aufgrund eines falschen Verständnisses von Wahrscheinlichkeiten, resultieren, so ist auch damit noch kein hinreichendes Argument für Nudges des Typs II formuliert. Insofern es sich hier um *intransparente* Maßnahmen der Beeinflussung handelt, gilt es im Sinne der Gewährleistung von Autonomie zu eruerien, inwiefern Individuen durch angemessene Informationen und edukative Maßnahmen instandgesetzt werden können, eigenständig die jeweils angemessene Heuristik zu wählen und somit bessere Entscheidungen zu treffen (Rachlinski 2003, Gigerenzer 2014, 2015). Im Falle des Umgangs mit Wahrscheinlichkeiten etwa hat sich gezeigt, dass die Darstellung des Entscheidungsproblems in Form von Häufigkeiten (zum Beispiel 3 von 10) anstatt subjektiver Wahrscheinlichkeiten (30 Prozent) Fehler minimiert (Rachlinski 2003: 1208). Insofern diese alternative Darstellung selbst als Nudge bezeichnet werden kann, können bei entsprechenden Entscheidungsproblemen edukative Nudges Abhilfe schaffen, die Lernprozesse seitens der Individuen ermöglichen (Sunstein 2016b: 38).

3.3 Nudge III und IV: Mangelnde Selbstkontrolle

In die Kategorie III fallen Nudges, die zwar eine mehr oder weniger automatische Verhaltensänderung bewirken, die aber von der beeinflussten Person als Einflussfaktoren erkannt werden können (Hansen und Jespersen 2013: 21). Dazu zählen etwa die Verwendung von Signalfarben, die Änderung der Standardeinstellung des Druckers oder auch von Sparplänen und Betriebsrenten (Thaler und Sunstein 2011: 148 ff.). Während diese Maßnahmen unmittelbar auf das automatische System abzielen und für ihre Wirkungsweise nicht auf reflexive Prozesse angewiesen sind, erlauben sie durch ihre relativ einfache Erkennbarkeit (Transparenz) die bewusste Reflektion und entsprechende Reaktion seitens des Individuums; ob das Papier etwa einseitig oder zweiseitig bedruckt wird oder ob die Gehaltsabrechnung den Abzug einer Rentenzahlung aufweist, ist leicht feststellbar. Zur Kategorie IV schließlich gehören Nudges, die im Gegensatz dazu schwierig für das Individuum erkennbar sind, da sie weder transparent sind noch auf das reflexive System wirken. Dies ist etwa der Fall beim Einfluss der Tellergröße auf die Kalorienannahme oder wenn Menschen von Geburt an, das heißt ohne explizite Einwilligung Organspender sind.

Diese beiden Typen des Nudges sind insofern problematisch, als sich die unmittelbare automatische Reaktion nicht kontrollieren lässt, das reflexive System also nicht beteiligt wird (Typ III) oder sie nicht explizit bemerkt werden (Typ IV). Dieses Legitimationsproblem lässt sich jedoch unter bestimmten Bedingungen lösen; dann nämlich, wenn Nudges als *Mittel der vernünftigen Selbstbindung* begriffen werden, die weiterhin autonome Entscheidungen zulassen. Unter der Voraussetzung, dass zum Beispiel die Verringerung von Kalorien und das Spenden von Organen von den Individuen grundsätzlich erwünscht sind, lassen sich sowohl Kategorie III als auch IV auf Entscheidungen beziehen, die Inkonsistenzen zwischen aktuellen und langfristigen Präferenzen widerspiegeln und vom betreffenden Individuum selbst als nicht in seine übergeordnete Präferenzstruktur oder seinen Lebensplan passend bewertet werden. In beiden Kategorien geht es somit um Probleme der Selbstkontrolle, deren nachteilige Wirkung auf das Individuum – im Gegensatz zu anderen Irrationalitäten – in der Literatur gut dokumentiert ist (Offer 2006, Michel 2015).

Um die Risiken fehlender Reflexion und mangelnder Legitimation auszuschließen, ist es angesichts des Umstands, dass in beiden Fällen Probleme der Selbstkontrolle vorliegen, sinnvoll, Nudges als *Mittel der vernünftigen Selbstbindung* zu begreifen, die weiterhin eigenverantwortliche Handlungen zulassen (Heidbrink 2015, 2018). Die Voraussetzung für vernünftige Selbstbindung

besteht darin, dass die verwendeten Nudges transparent und als Instrumente der Verhaltenslenkung erkennbar sind, wie im Falle von Nudges vom Typ I. Zudem müssen sie den betroffenen Individuen, etwa durch deliberative Zustimmungsverfahren, autonome Entscheidungen ermöglichen (im Falle von Nudges vom Typ III und IV). Aus dieser Perspektive betrachtet sind unterstützende politische Maßnahmen keine Erfindung des libertären Paternalismus, sondern knüpfen an ältere moralphilosophische und demokratietheoretische Debatten zum Umgang mit Rationalitätsanomalien an (Offe 1989, Elster 1990, Schaal und Ritzi 2009).

Die Verwendung von Nudges als Mittel der vernünftigen Selbstbindung stellt einen wichtigen Ansatzpunkt dar, um zwischen legitimen und illegitimen Nudges differenzieren zu können. Für politische Verfahren der Verhaltenslenkung muss gewährleistet sein, dass die verwendeten Nudges sowohl erkennbar sind und reflexiv bewertet werden können als auch, dass die Individuen dem Ziel und den Maßnahmen *tatsächlich* zustimmen können. Um dies sicherzustellen reichen weder ambitionierte „Neujahrsaussagen“, also öffentlich proklamierte, aber unrealistische oder nicht durchdachte Vorsätze, noch die Beobachtung anscheinend inkonsistenten Verhaltens aus. Es schließt sich die Frage an, wie dieses Erfordernis praktisch umgesetzt werden kann, insbesondere ohne die Rechte derer zu verletzen, die kein Interesse an entsprechender Selbstbindung haben. Neben lokalen und privaten Anbietern von Mechanismen der Selbstbindung (zum Beispiel stickk.com, eine Internetseite, auf der sich Nutzer öffentlich und gegen einen bestimmten Einsatz auf selbstgesetzte Ziele verpflichten können, Thaler und Sunstein 2011: 300 f.) spielt hier die Kommunikation entsprechender Maßnahmen eine wichtige Rolle. Denn nur wenn die Betroffenen informiert und in die Gestaltung der Maßnahmen des Nudging einbezogen werden, können sie entsprechende Maßnahmen erkennen und sich der vom „Nudgenden“ intendierten Verhaltensbeeinflussung bewusst widersetzen (Bovens 2009, Hanson und Jespersen 2013).

4 Fazit

Wir haben dafür argumentiert, dass eine normative Begründung des libertären Paternalismus sich nicht primär auf Abweichungen individueller Entscheidungen vom Rational Choice Ideal stützen kann. Dabei ist nämlich weder klar, ob die festgestellten Abweichungen überhaupt Nachteile für das Individuum zur Folge haben, noch ist ausgemacht, dass sie am besten durch Nudging behoben werden können. Auf Basis dieser Überlegung haben wir die Nudges nach den Kriterien der Transparenz und der Wirkungsweise auf die zwei Denksysteme in einer Matrix kategorisiert. Daran lässt sich zeigen, dass Nudges dann unproblematisch sind, wenn sie sich als transparente Mittel der vernünftigen Selbstbindung begreifen lassen. Dies ist der Fall bei Nudges vom Typ I, die aus Unachtsamkeit und Vergesslichkeit resultieren, sowie bei den Typen III und IV, sofern den Maßnahmen zugestimmt worden ist und diese auf von den Individuen selbst als problematisch bewertete Inkonsistenzen zwischen konkretem Handeln einerseits und übergeordneten Präferenzen andererseits abstellen. Leitend ist dabei die Überlegung, dass Nudges weiterhin autonomes Handeln zulassen. Bei Nudges vom Typ II ist dies der Fall, sofern sie eine edukative Funktion haben.

Aus den Ergebnissen lassen sich für die weitere Forschung Schlussfolgerungen für die Verwendung von Nudging in der politischen Planung, Lenkung und Gestaltung ziehen. Nudges können als *Mandate* aufgefasst werden (Frank 2008), die an politische Steuerungsinstanzen vergeben werden, um eine vernünftige Selbstbindung gesellschaftlicher Akteure zu ermöglichen. Die mög-

liche Einschränkung individueller Freiheiten steht dabei in keinem Widerspruch zur Norm der Autonomie, solange die Maßnahmen transparent sind, auf deliberativen Verfahren der Selbstbindung beruhen und Lernprozesse ermöglichen (Heidbrink 2018).

Literaturverzeichnis

- Baldwin, Robert (2014): From regulation to behavior change: giving nudge the third degree. *The Modern Law Review*, 77 (6), 831–857.
- Berg, Nathan und Gerd Gigerenzer (2010): As-if behavioral economics: neoclassical economics in disguise? *History of Economic Ideas* 17 (1), 133–165.
- Blumenthal-Barby, Jennifer S. (2012): Between reason and coercion: ethically permissible influence in health care and health policy contexts. *Kennedy Institute of Ethics Journal* 22 (4), 345–366.
- Bovens, Luc (2009): The ethics of nudge. In: Till Grüne-Yanoff und Sven Ove Hansson (Hrsg.): *Preference change: Approaches from philosophy, economics and psychology*. Berlin et al., Springer, 207–219.
- Elster, John (1990): *Ulysses and the Sirens: Studies in rationality and irrationality*. Cambridge et al., CUP.
- Frank, Robert H. (2008): Richard H. Thaler and Cass R. Sunstein: *Nudge. Improving decisions about health, wealth, and happiness*. *Ethics*, 119 (1), 202–208.
- Frankfurt, Harry (2001): *Freiheit und Selbstbestimmung*, Berlin, Akademie.
- Gigerenzer, Gerd (2014): *Risk Savvy. How to make good decisions*. London et al., Penguin.
- Gigerenzer, Gerd (2015): On the supposed evidence for libertarian paternalism. *Review of Philosophical Psychology*, 6, 361–383.
- Glaeser, Edward (2006): Paternalism and psychology. *The University of Chicago Law Review*, 73, 133–156.
- Grüne-Yanoff, Till (2012): Old wine in new casks: Libertarian paternalism still violates liberal principles. *Social Choice and Welfare*, 38, 635–645.
- Hanna, Jason (2015): Libertarian paternalism, manipulation and the shaping of preferences. *Social Theory and Practice*, 41 (4), 618–643.
- Hansen, Pelle Guldborg (2016): The definition of nudge and libertarian paternalism: does the hand fit the glove? *European Journal of Risk Regulation*, 7 (1), 155–174.
- Hansen, Pelle Guldborg und Andreas Maaloe Jespersen (2013): Nudge and the manipulation of choice. A framework for the responsible use of the nudge approach to behavior change in public policy. *European Journal of Risk Regulation*, 4 (1), 3–28.
- Hausman, Daniel M. (2012): *Preference, Value, Choice, and Welfare*. Cambridge et al., CUP.
- Hausman, Daniel M. und Brynn Welch (2010): Debate: To nudge or not to nudge. *The Journal of Political Philosophy*, 18 (1), 123–136.
- Heidbrink, Ludger (2015): Libertarian paternalism, sustainable self-binding and bounded freedom. In: Dieter Birnbacher und May Thorseth (Hrsg.): *The politics of sustainability. Philosophical perspectives*. London/New York, Routledge, 173–194.
- Heidbrink, Ludger (2018): Selbstbindung als wirtschaftsethisches Problem. Zur Rolle des libertären Paternalismus bei der Förderung von Wohlfahrtsentscheidungen. In: Arne Manzeschke (Hrsg.): *Evangelische Wirtschaftsethik – wohin? Perspektiven und Positionen*. Münster, LIT, 217–238.

- Infante, Gerardo, Guilhem Lecouteux und Robert Sugden (2016): Preference purification and the inner rational agent: A critique of the conventional wisdom of behavioral welfare economics. *Journal of Economic Methodology*, 23 (1), 1–25.
- Kahneman, Daniel (2011): Thinking fast and slow. London, Allen Lane.
- Kemmerer, Alexandra, Christoph Möllers, Maximilian Steinbeis und Gerald Wagner (Hrsg.) (2006): Choice architecture in democracies. Exploring the legitimacy of nudging. Baden-Baden, Nomos.
- Klonschinski, Andrea und Joachim Wündisch (2016): Präferenzen, Wohlergehen und Rationalität – Zu den begrifflichen Grundlagen des libertären Paternalismus und ihren Konsequenzen für seine Legitimierbarkeit. *Zeitschrift für Praktische Philosophie*, 3 (1), 599–632.
- Lusk, Jayson L. (2014): Are you smart enough to know what to eat? A critique of behavioral economics as justification for regulation. *European Review of Agricultural Economics*, 41 (3), 355–373.
- McQuillin, Ben und Robert Sugden (2012): Reconciling normative and behavioral economics: the problems to be solved. *Social Choice and Welfare*, 38, 553–567.
- Mischel, Walter (2015): The Marshmallow Test. Understanding self-control and how to master it. London, Corgi.
- Neumann, Robert (2013): Libertärer Paternalismus. Theorie und Empirie staatlicher Entscheidungsarchitektur. Stuttgart, Mohr Siebeck.
- Offe, Claus (1989): Fessel und Bremse. Moralische und institutionelle Aspekte „intelligenter Selbstbeschränkung“. In: Axel Honneth et al. (Hrsg.): Zwischenbetrachtungen. Im Prozeß der Aufklärung. Jürgen Habermas zum 60. Geburtstag, Frankfurt a. M., Suhrkamp, 739–774.
- Offer, Avner (2006): The challenge of affluence. Oxford et al., OUP.
- Rachlinski, Jeffrey J. (2003): The uncertain psychological case for paternalism. *Northwestern University Law Review*, 97 (3), 1165–1225.
- Rebonato, Riccardo (2012): Taking liberties. A critical examination of libertarian paternalism. Basingstoke, Palgrave Macmillan.
- Schaal, Gary S. und Claudia Ritzi (2009): Rationale Selbstbindung und die Qualität politischer Entscheidungen – liberale und deliberative Perspektiven. In: Gary S. Schaal (Hrsg.): Techniken rationaler Selbstbindung, Berlin, LIT, 55–95.
- Sugden, Robert (2008): Why incoherent preferences do not justify paternalism. *Constitutional Political Economy*, 19, 226–248.
- Sugden, Robert (2009): On nudging: a review of nudge: Improving decisions about health, health, and happiness by Richard H. Thaler and Cass R. Sunstein. *International Journal of the Economics of Business*, 16 (3), 365–373.
- Sunstein, Cass R. (2013): Simpler. The future of government. New York, Simon and Schuster.
- Sunstein, Cass R. (2015): Choosing not to choose. Understanding the value of choice. Oxford et al., OUP.
- Sunstein, Cass R. (2016a): The ethics of influence. Government in the age of behavioral science. New York, CUP.
- Sunstein, Cass R. (2016b): The ethics of choice architecture. In: Alexandra Kemmerer et al. (Hrsg.) 21–74.
- Sunstein, Cass R. und Richard H. Thaler (2003): Libertarian paternalism is not an oxymoron. *The University of Chicago Law Review*, 70, 1166–1187.

- Thaler, Richard R. (2015): *Misbehaving. The making of behavioral economics.* New York/London, W. W. Norton and Company.
- Thaler, Richard H. und Cass R. Sunstein (2003): Libertarian paternalism. *American Economic Review Papers and Proceedings*, 93 (2), 175–179.
- Thaler, Richard H. und Cass R. Sunstein (2009): *Nudge. Improving decisions about health, wealth, and happiness.* London et al., Penguin Books.
- Thaler, Richard H. und Cass R. Sunstein (2011): *Nudge. Wie man kluge Entscheidungen anstößt.* Berlin, Ullstein.
- Van Aaken, Anne (2016): Constitutional limits to nudging: A proportionality assessment. In: Kemmerer et al. (Hrsg.), 161–195.
- White, Mark D. (2013): *The manipulation of choice. Ethics and libertarian paternalism.* New York, Palgrave.
- Wilkinson, T. Martin (2013): Nudging and manipulation. *Political Studies*, 16, 341–355.