UNIVERSITY OF HAMBURG

DOCTORAL THESIS

Environmental Policy: How context affects behavior

Author:
Hendrik BRUNS
Born in:
Celle

A thesis submitted in partial fulfillment of the requirements for the degree of Doktor der Wirtschafts- und Sozialwissenschaften

at the

Faculty of Business, Economics and Social Sciences
Department of Socioeconomics
Chair of Ecological Economics

(according to PromO from August 24, 2010)

Hamburg, May 8, 2019 (print year)

Head of committee:
Prof. Dr. Miriam BEBLO

First supervisor:
Prof. Dr. Grischa PERINO

Second supervisor:
Prof. Dr. Andreas LANGE

Date of defense: 30.08.2018

"You look at where you're going and where you are and it never makes sense, but then you look back at where you've been and a pattern seems to emerge."

Robert M. Pirsig

UNIVERSITY OF HAMBURG

Abstract

Faculty of Business, Economics and Social Sciences
Department of Socioeconomics

Doktor der Wirtschafts- und Sozialwissenschaften

Environmental Policy:
How context affects behavior

by Hendrik BRUNS

This thesis focuses on two important aspects of climate change. First, it deals with recent behavioral instruments to induce individual climate protection, as well as their behavioral underpinnings. Second, it investigates possible factors supporting the emergence of skepticism towards climate change in the light of increasing online news consumption. All of the studies presented here draw insights based on lab- and field experimental data and test causal hypotheses. Central findings are: (1) Neither does transparency on the potential influence and purpose of a pro-environmental default reduce its effectiveness, nor do different aspects of psychological reactance change that. (2) Defaults can have detrimental effects on highly intrinsically motivated contributions to climate protection, relative to recommendations and mandatory minimum contributions, although interaction with source information appears to be negligible. (3) Proclaimed doubt in the independence of news media does not appear to negatively affect consumers' trust in news media- and scientific sources reporting on recent events in climate change. Findings contribute to a better understanding of how pro-environmental nudges work, their effectiveness in relation to conventional instruments, and of the role of proclaimed media distrust on public trust in media and scientific sources reporting on climate change. The former is central to an improved understanding of behavioral instruments that become increasingly important worldwide as governmental policy-tools. The latter is essential for a better understanding of contemporary online environments and their impact on the emergence of fake news, echo chambers, and filter bubbles.

UNIVERSITY OF HAMBURG

Zusammenfassung

Faculty of Business, Economics and Social Sciences
Department of Socioeconomics

Doktor der Wirtschafts- und Sozialwissenschaften

Environmental Policy:
How context affects behavior

von Hendrik BRUNS

Diese Thesis widmet sich zwei zentralen Aspekten des Klimawandels. Erstens beschäftigt sie sich sowohl mit neueren Verhaltensinstrumenten zur Motivation individuellen Klimaschutzverhaltens, als auch mit deren Verhaltensgrundlagen. Zweitens untersucht sie mögliche Einflussfaktoren auf die Entstehung von Skeptizismus gegenüber Klimawandel, insbesondere vor dem Hintergrund des ansteigenden online Nachrichtenkonsums. Alle der hier vorgestellten Studien ziehen Schlussfolgerungen auf Basis labor- und feldexperimenteller Daten und testen kausale Hypothesen. Wesentliche Ergebnisse sind: (1) Weder beeinflusst Transparenz in Bezug auf potentielle Effekte und Absichten eines umweltfreundlichen Standardwertes dessen Effekt, noch wird dies durch Aspekte psychologischer Reaktanz verändert. (2) Im Vergleich zu Empfehlungen und verpflichtenden Mindestbeiträgen können Standardwerte nachteilige Effekte auf Klimaschutzbeiträge mit hoher intrinsischer Motivation haben. Die Interaktion mit Informationen zum Urheber der Interventionen erscheint vernachlässigbar zu sein. (3) Verkündetes Misstrauen in die Unabhängigkeit der Medien scheint weder einen negativen Einfluss auf das Vertrauen von Konsumenten in Nachrichtenmedien, noch auf das Vertrauen in wissenschaftliche Quellen, die zum Klimawandel berichten, zu haben. Die Ergebnisse tragen zu einem besseren wissenschaftlichen Verständnis der Wirkungsweise umweltfreundlicher "Nudges", ihrer Effektivität im Vergleich zu konventionellen Instrumenten, und der Rolle verkündeten Medienmisstrauens auf das öffentliche Vertrauen in Medien- und wissenschaftliche Beiträge zum Klimawandel bei. Ersteres ist zentral um Verhaltensinstrumente, welche als Politikinstrumente weltweit an Bedeutung gewinnen, besser zu verstehen. Letzteres ist insbesondere für ein besseres Verständnis von gegenwärtigen Online-Umgebungen und ihrem Einfluss auf die Entstehung von Falschmeldungen, Echokammern, und Filterblasen von Bedeutung.

Acknowledgements

I am deeply thankful to my supervisor Prof. Dr. Grischa Perino for his patience, support, and advice throughout my dissertation. Special thanks go out to Dr. Claudia Schwirplies for patient and honest support, as well as Maximilian Willner for insightful discussion and for keeping up the good social spirits in our office. I also want to thank my co-authors Dr. Elena Kantorowicz-Reznichenko, Dr. Marijane Luistro Jonsson, Katharina Klement, and Bilel Rahali for accompanying me to my first scientific publication. I am also grateful to the Graduate School of the Faculty of Business, Economics and Social Sciences for financial support, as well as the excellent support and training provided by the International Max-Planck Research School on Earth System Modelling of the Max-Planck Institute of Meteorology in Hamburg. I thank Prof Dr. Andreas Lange, Prof. Dr. Hermann Held for their professional supervision in my Advisory Panel Meetings, as well as Dr. Johannes Jarke and Jascha Tutt for fruitful discussions. Last but not least I thank my mother, as well as my best friends Bastian, Lennart, Nis, Hannes, Eike, Hannah, and Maike.

Contents

Abstract **v**

Zusammenfassung **vii**

Acknowledgements **ix**

List of Figures **xiii**

List of Tables **xv**

List of Abbreviations **xvii**

1 Introduction **1**

2 Can nudges be transparent and yet effective? **9**
2.1 Introduction 10
2.2 Conceptual framework and behavioral predictions 13
2.3 Experimental design 15
2.4 Results 18
2.4.1 Default effects 18
2.4.2 Influence of transparency on default effectiveness . . . 20
2.4.3 Psychological reactance and transparency 21
2.5 Discussion and conclusion 23

3 Point at, nudge, or push private provisions to a public good? Field experimental evidence for experts, politicians, and nobodies **29**
3.1 Introduction 30
3.2 Experiment 32
3.2.1 Experimental procedure and treatments 32
3.2.2 Outcome measures 34
3.2.3 Behavioral predictions 34
3.2.4 Participants 36
3.2.5 Statistical analyses and power analyses 37
3.3 Findings 38
3.3.1 Descriptive statistics 38
3.3.2 Finding 1: Effect of intervention type on contribution changes 39
3.3.3 Finding 2: Interaction of intervention type and intrinsic motivation 41
3.3.4 Finding 3: Effect of source type on contribution changes 44

3.3.5 Finding 4: Interaction of source type and intrinsic motivation . . . 46
3.3.6 Finding 5: Interaction of intervention- and source type 48
3.4 Discussion . . . 49
3.5 Conclusion . . . 50

4 Does proclaimed doubt in media spill over to doubt in science? A laboratory experiment in the context of climate change 53
4.1 Introduction . . . 54
4.2 Relevant literature and hypotheses . . . 55
4.3 Experiment . . . 58
4.3.1 Experimental design and manipulation . . . 58
4.3.2 Outcome measures . . . 59
4.3.3 Participants . . . 59
4.3.4 Procedure . . . 60
4.3.5 Statistical analyses . . . 60
4.4 Findings . . . 63
4.4.1 Descriptive statistics . . . 63
4.4.2 Treatment effects on primary outcomes . . . 63
4.4.3 Treatment effects on secondary outcomes . . . 67
4.5 Discussion and Conclusion . . . 68

5 Conclusion 71

Appendix A 75
A.1 Experimental design . . . 75
A.2 Statistical analyses . . . 79
A.2.1 Comparing subjects from Rotterdam and Hamburg . . . 82
A.3 Questionnaire . . . 86

Appendix B 89
B.1 Experimental design . . . 89
B.1.1 Experimental screens (translated) . . . 89
B.1.2 Instructions (translated) . . . 93
B.2 Statistical analyses . . . 95
B.2.1 Information on participant selection . . . 100
B.3 Questionnaire . . . 102
B.3.1 Pre-experimental questionnaire (Stage 1) . . . 102
B.3.2 Post-experimental questionnaire (Stage 2) . . . 104

Appendix C 105
C.1 Instructions and materials . . . 105
C.2 Statistical analyses . . . 110
C.3 Questionnaire . . . 115

Bibliography 117

List of Figures

2.1 Mean contributions per experimental group 20

3.1 Categorized contributions by experimental group. 39
3.2 Kernel probability densities of contribution changes (> 35) by type of intervention 40
3.3 Intervention effects on contribution changes (> 35) relative to control conditional on baseline contributions 42
3.4 Kernel probability densities of contribution changes (> 35) by type of source 45
3.5 Source effects on contribution changes (> 35) relative to no source information conditional on baseline contributions (> 35) for different intervention types 47
3.6 Predicted values of contribution changes (> 35) for source type conditional on intervention type 48

A.1 Experimental screen for Control 77
A.2 Experimental screen for Default + transparency 78
A.3 Distribution of contributions 79
A.4 Default and transparency effects on contributions for different base-categories 80
A.5 Default and transparency effects on perceived Threat to freedom 81
A.6 Default and transparency effects on Anger 81
A.7 Mean contributions by experimental group and location . . . 84
A.8 Effect sizes by location and for aggregated data 85
A.9 Coefficients from tobit model by location and for aggregated data 85

B.1 Test question 89
B.2 Information about the source of the respective intervention (political) 90
B.3 Decision screen for baseline contribution 90
B.4 Decision screen for recommendation intervention (with source information) 91
B.5 Decision screen for default intervention (with source information) 91
B.6 Decision screen for restriction intervention (with source information) 92

C.1 Two prior distributions, and a Gauss distribution for reference 110

C.2 Trends related to usage of the term 'fake news' by Twitter account of Donald Trump 111
C.3 Violin plots of trust in media and science 112
C.4 Violin plots of perceptions of trust, expertise, and content credibility in media and science 113

List of Tables

2.1 Experimental design 17
2.2 Descriptive statistics of all outcome variables to assess the default effect 19
2.3 Descriptive statistics of covariates 21
2.4 Stepwise Tobit-models with and without interaction term . . 22
2.5 Ordered logistic model of state reactance 24

3.1 Experimental design 33
3.2 Descriptive statistics of contributions by experimental group . 38

4.1 Trustworthiness by source and treatment 63
4.2 Source and content credibility by source and treatment 64
4.3 Ordinal logistic regression of media and science trustworthiness on treatment and covariates, and treatment-order interaction 64
4.4 Bayesian logistic regression of media trustworthiness on treatment and covariates 65
4.5 Bayesian logistic regression of science trustworthiness on treatment and covariates 66
4.6 OLS-regression of media trustworthiness, expertise, and content credibility on treatment 68
4.7 OLS-regression of science trustworthiness, expertise, and content credibility on treatment 68

A.1 P-values for pairwise MW tests of Contribution 79
A.2 Descriptive statistics of all outcome variables by experimental group and location 83
A.3 Descriptive statistics of covariates by experimental group and location 84

B.1 Descriptive statistics of reactance proneness, environmental orientation, political orientation, and attitudes towards license retirement 95
B.2 Descriptive statistics of gender, education, and income 96
B.3 P-values from pairwise Mann-Whitney tests for equality of contribution changes between experimental groups (excluding restriction treatments) 97
B.4 P-values from pairwise Mann-Whitney tests for equality of contribution changes > 35 between experimental groups . . . 97

B.5 Linear regression models of intervention type and interaction with baseline contributions on contribution changes ($>$ 35) without source information 98
B.6 Linear regression models of source type and interaction with baseline contributions ($>$ 35) on contribution changes ($>$ 35), conditional on intervention types 98
B.7 Linear regression models of intervention type and source type, including interaction, on contribution changes ($>$ 35) 99
B.8 Linear regression model for carryover effects 99

C.1 Ordinal logistic regression of media and science trustworthiness on treatment and covariates 112
C.2 OLS-regression of media trustworthiness, expertise, and content credibility on treatment and covariates 114
C.3 OLS-regression of science trustworthiness, expertise, and content credibility on treatment and covariates 114

List of Abbreviations

CO_2	**C**arbon Di**o**xide
EUETS	**E**uropean **U**nion **E**missions **T**rading **S**cheme
GHG	**G**reen**h**ouse **G**as
MWU	**M**ann **W**hitney **U** test
NOAA	**N**ational **O**ceanic and **A**tmospheric **A**dministration
NUTS	**N**o-**U**-**T**urn-**S**ampler
OLS	**O**rdinary **L**east **S**quares
RLD	**R**andom **L**ottery **D**esign
SIF	**S**upposedly **I**rrelevant **F**actor
UNFCCC	**U**nited **N**ations **F**ramework **C**onvention on **C**limate **C**hange
US	**U**nited **S**tates of America
QRP	**Q**uestionable **R**esearch **P**ractice

Dedicated to my cat.

1 Introduction

Finding solutions to climate change is one of the biggest inter- and intragenerational challenges for societies on a global scale. Temperature-, as well as a variety of co-occurring changes, e.g. shifts in precipitation patterns, sea level rise, glacier meltdown, intensification of and changes in the oscillation of extreme weather events, will have dire ecological, as well as social and economic consequences on macro- and micro scales. These intertemporal changes will affect nation states and people irrespective of their causal role in climate change, and regardless of their economic, social, and ecological means to mitigate, or adapt to these changes (IPCC, 2014). In other words, those responsible for climate change are neither necessarily those that suffer the most, nor are they automatically viewing themselves as the ones most responsible to mitigate the consequences for everyone. This situation necessitates that strategies to tackle climate change be developed - and that these instruments will be accessible to most, if not all actors that face negative consequences - in the present or future.

This thesis focuses on two important aspects of climate change. First, it deals with current instruments to induce individual climate protection, as well as with their behavioral foundations. Second, it investigates possible factors underlying climate change skepticism developing in contexts of online news consumption. All of the studies presented here rely on experimental data and test causal hypotheses.

Political strategies to tackle climate change have ambitious goals. In December 2015, member states of the United Nations Framework Convention on Climate Change (UNFCCC) adopted the Paris Agreement. The agreement postulates the aim to limit the average global temperature increase to 2 °C, including efforts to limit the increase to 1.5 °C above pre-industrial levels (UNFCCC, 2015). The ambitiousness of this target becomes apparent when realizing that integrated assessment models assume the large-scale use of technologies to realize *negative* emissions in order to reach this target in the allotted time (Anderson and Peters, 2016).

Still, an important strategy to tackle climate change is mitigation, incorporating strategies to motivate reductions of greenhouse-gases (GHG) by individuals. Generally, strategies to cope with climate change broadly fall into the categories a) mitigation, b) adaptation, and c) climate engineering. Of these, mitigation aims at realizing the temperature-limit goals postulated in the Paris Agreement primarily by means of reducing emissions of carbon dioxide (CO_2) to the atmosphere. In order to reach the ambitious goals set in Paris, mitigation strategies, as well as carbon sequestration to achieve negative emissions, need to be applied soon and on large scales (Anderson

and Peters, 2016). This includes national, sub-national, but also non-state actions, such as individual lifestyle changes, especially from the high-carbon-emitting individuals estimated to produce nearly 50 % of emissions (Wynes and Nicholas, 2017; Rogelj et al., 2016; van Vuuren et al., 2018).

Voluntary individual carbon mitigation as a contribution to climate protection is subject to intensive scientific investigation, especially in the fields of economics and psychology. Conceptually, individual behavior such as lifestyle changes, e.g. reducing electricity consumption, living car-free, avoiding airplane travel, or eating vegetarian (Wynes and Nicholas, 2017), can all be understood as contributions to a (large) public good. Formally, the supply of a public good resembles a dilemma-situation, in which there is a lack of incentives for rational actors to forfeit individual gains for the benefit of others. Therefore, rational actors would not contribute to public goods, i.e. they would not reduce their CO_2 emissions, constituting a market failure resulting in climate change (Stern, 2008). However, numerous empirical and experimental findings have shown that individuals contribute to public goods, specifically to climate protection, e.g. by means of voluntary direct and indirect GHG mitigation (e.g. Fehr and Gächter, 2000; Fehr and Fischbacher, 2003; Gintis et al., 2005; Lange et al., 2014). Consequently, various motives, respectively preferences such as fairness, reciprocity, conditional cooperation, altruism, warm-glow, intrinsic motivation, social norms, image motivation, or inequality aversion, have been invoked to rationalize observed contributions (e.g. Sugden, 1984; Andreoni, 1988; Fehr and Schmidt, 1999; Bénabou and Tirole, 2006).

Based on this, environmental economics and environmental psychology have come up with various instruments to motivate public good contributions, esp. by means of inducing individuals to reduce their GHG emissions. The most prominent instruments can be subsumed under the categories a) information provision/ education, b) economic incentives (prices and quantities), and c) mandates/ bans. As such, these instruments range from informing people about possible behaviors that reduce their emissions, over providing monetary incentives for climate-friendly, and disincentives for climate-damaging behavior, to mandates and bans of these. However, all of these instruments have been exposed to different forms of criticism. For example, economic interventions, especially environmental taxes and subsidies, have been criticized for their potential to negatively influence peoples' intrinsic motivation to protect the environment (e.g. Frey, 1992; Cardenas et al., 2000; Goeschl and Perino, 2012; Perino et al., 2014; Rode et al., 2015). So have bans and mandates, which bear the same potential for negative side-effects (Falk and Kosfeld, 2006). This negative interaction of interventions with the intrinsic motivation of individuals is labeled motivation crowding-out and limits their potential of mitigating individual GHG emissions. Similarly, the provision of information to foster pro-environmental attitudes, expected to induce respective behavior change, has repeatedly been shown to be limited by the attitude-behavior-gap (also: value-action gap) (Kollmuss and Agyeman, 2002). In other words, being informed that the environment needs protection, or that individual reductions in GHG emissions are needed in order

to reduce global warming, does not automatically result in the respective behavior, even when individuals share these views.

One concept to explain the attitude-behavior-gap and motivation crowding comes from the intersection of economics and cognitive psychology: bounded rationality. This concept can be seen as the core of behavioral economics and has thus also become central to behavioral environmental economics (Croson and Treich, 2014). Bounded rationality is a model of human decision making brought forward as an alternative to the assumption of perfect rationality in neoclassical economics, i.e. the *homo oeconomicus*. Boundedly rational decision making is characterized by the use of behavioral heuristics, i.e. rules-of-thumb, and the influence of behavioral anomalies, both resulting from limits to cognitive capacity, willpower and self-interest characterizing "real" humans (Simon, 1955; Simon, 1990; Thaler, 2016). Thus, models of bounded rationality, contrary to their counterparts assuming perfect rationality, incorporate behavioral regularities that have been empirically found to systematically affect human judgment, decision making, and ultimately behavior. Examples of such empirical regularities are cognitive myopia, loss aversion, hyperbolic time discounting, distaste of risk and ambiguity, the effect of small probabilities, affect-based, or rule- and role-based decisions (see Weber and Johnson, 2012).

The incorporation of bounded rationality into (behavioral environmental) economics resulted in new instruments of behavioral change: nudges. Behavioral economics took note of the systematic departures of behavior from perfect rationality and included them in their models. Picking up pace in the 1980s, its emergence resulted not only in the creation of empirically-founded theories and models of economic decision making, abandoning the assumption of rationality, but also in complementary, or even substituting interventions to induce behavioral change, such as motivating individuals to behave pro-environmentally (Gowdy, 2008; Shogren and Taylor, 2008; Croson and Treich, 2014). Behaviorally informed instruments rely on the influence of "supposedly irrelevant factors" (SIFs). SIFs had hitherto been "assumed away" based on the argument that they were not relevant in market contexts and that models should assume that market actors behaved "as if" they were perfectly rational and thus insensitive to SIFs (Thaler, 2016). Interventions using SIFs are often called nudges and are usually defined as interventions that alter behavior without using financial incentives, or limiting choice options (Thaler and Sunstein, 2008). Although this is the core of their definition, additional characteristics are that these instruments a) attempt to change behavior directly, instead of indirectly via attitudes, b) primarily work subconsciously, and c) exploit, or rely on, the bounded rationality characterizing human decision making by changing SIFs. An example of a SIF is the default selection of one of several options between which a decision has to be made. A rational actor would choose an option according to his preferences, maximizing his private utility, irrespective of the chosen default. A boundedly rational actor, however, might consider the default option first, and then evaluate all other options in relation to it, potentially increasing the chance of choosing the default value (Dhingra et al., 2012). Very recently, nudges

have been applied to an increasing degree by national governments, as well as NGOs in order to induce behavioral change. Areas of application range from financial security, over education, energy-saving, health, and job training, to program integrity and compliance, as well as home affairs. One of their advantages is that they have been relatively cost-efficient (Benartzi et al., 2017).

Naturally, while nudges gained a lot of political, as well as scientific attention in the years following their emergence, they have also been subject to various forms of criticism. While these criticisms are also philosophical (Rebonato, 2014), or theoretical (Grüne-Yanoff and Hertwig, 2015), they are often empirical, i.e. questioning their effectiveness, efficiency, and ethicality (Sunstein, 2018). The first two papers presented in this thesis contribute to a better understanding of the ethicality of nudges, as well as their effectiveness in relation to conventional instruments. As a specific nudge, default values for individual contributions to climate protection are investigated. The latter is operationalized as monetary payments to buy and retire carbon emission allowances of the European Union Emissions Trading Scheme (EUETS). The first paper provides evidence on the question whether nudges work if decision makers are made aware that they are being nudged, and whether individual traits and perceptions moderate its effect. The second paper evaluates the relative effectiveness of nudges compared to conventional policy-instruments. It focuses on their potential to have unintended side-effects, as well as on the influence of information provided on the actor responsible for their implementation.

The first paper, titled *"Can nudges be transparent and yet effective?"*, builds on recent research suggesting that nudges potentially motivate people to (sub-) consciously behave opposite to the intervention's goal (e.g Costa and Kahn, 2013; Haggag and Paci, 2014; Hedlin and Sunstein, 2016; Arad and Rubinstein, 2017). Such side-effects can be a means of protesting against interventions that aim to change behavior or attitudes and are perceived by decision makers as a limit to their behavioral freedom, or autonomy. The mental and behavioral state causing such a reaction is known in the psychological literature as psychological reactance (Brehm, 1966). Another critique, which, for a long time, based primarily on anecdotal evidence deduced from the theoretical foundation of the nudge paradigm, claims that nudges only work if people are not aware of being nudged. Replicating and extending the work of Loewenstein et al. (2015) and other researchers (Kroese et al., 2016; Steffel et al., 2016; Bang et al., 2018), this thesis does not find evidence that nudges "work best in the dark" (Bovens, 2009, p. 217), or that psychological reactance accounts for variance in behavioral responses to transparent nudges. The latter finding is supported by findings of Goswami and Urminsky (2016).

The second paper has the title *"Point at, nudge, or push private provisions to a public good? Field experimental evidence for experts, politicians, and nobodies"* and starts from the empirically founded premise that behavioral autonomy has intrinsic, and not just instrumental value (Falk and Kosfeld, 2006; Bartling et al., 2014). Applying this to the recent application of nudges as governmental

policies to affect individual decision making, the paper investigates whether defaults interact differently with intrinsic motivation (if they interact at all), compared to recommendations and mandated contributions to climate protection. Thus, the paper extends the research of motivation crowding, which has exclusively been discussed in relation to economic instruments, to the nudge paradigm. Additionally, the paper builds on recent research that provides evidence on the potential role of regulator information on the performance of these interventions, especially defaults (e.g. Brown and Krishna, 2004; Altmann et al., 2015; Tannenbaum et al., 2017). Central results indicate that neither intervention increases average contributions to climate protection. Additionally, it shows that low default values can crowd out highly intrinsically motivated people, and that information on the source responsible for the implementation of any intervention does not change how decision makers react to these interventions.

The third paper, titled *"Does proclaimed doubt in media spill over to doubt in science? A laboratory experiment in the context of climate change"*, focuses on another challenge important in the context of the Paris Agreement, and climate change more generally: climate change skepticism and denial. Skepticism and denial concerning anthropogenic climate change can be exemplified by the withdrawal of the United States under the Trump Administration from the Paris Agreement in 2017 (Shear, 2017). Although most likely not its only explanation, the withdrawal is in line with climate change skepticism, and is often described as difficult to comprehend from a scientific perspective, because the causal role of humans on climate change is mostly agreed upon among scientists (Dunlap, 2013; Cook et al., 2016). Nevertheless, skepticism exists, and it has various facets, which can be political, societal, public, private, as well as (pseudo-) scientific. Consequently, climate change skepticism has become the subject of scientific investigation, resulting in multiple theories attempting to explain climate skepticism not primarily as a consequence of lacking intelligence, but rather as a result of various types of motivated reasoning (e.g. McCright and Dunlap, 2011; Kahan et al., 2012; Campbell and Kay, 2014). Very generally, motivated reasoning, or motivated cognition, describes a process of making inferences not primarily based on the objective and unbiased application of contextual information, but rather based on using information in order to arrive at a pre-set goal (Kunda, 1990). Such a goal could be to maintain, or strengthen one's cultural identity, i.e. identification with a (cultural) group, keeping one from holding beliefs that could alienate someone from his or her peer group (Kahan et al., 2007; Kahan, 2017; Kahan et al., 2017). As a specific example, it has been proposed that climate change deniers oppose the existence of climate change because they do not like its suggested solutions The underlying reason is that proposed instruments often rely on governments intervening in the market, which tends to disagree with the political beliefs held among Republicans, but not among Democrats. This is consistent with the empirical observation that Republicans are more likely to be skeptical of climate change, compared to Democrats (Campbell and Kay, 2014). Such forms of motivated cognition are characterized by a biased consumption of information, in order to avoid contradicting with views

and beliefs shared with peers, and thus lowering the risk of being expelled from the group (Kahan, 2017).

Motivated skepticism of climate change is fueled by the rapid growth of online environments used for public discussions. Like for other topics, debates on climate change are increasingly happening online, including the comment sections of online newspaper outlets, blogs, as well as social networks like Facebook and Twitter (e.g. Dunlap, 2013; Newman, 2017; Walter et al., 2018). These environments facilitate, and almost appear to consolidate the emergence of filter bubbles, fake news, and echo chambers (Lazer et al., 2018). These concepts, as well as their role to exert social influence have recently become the focus of scientific investigation (e.g. Allcott and Gentzkow, 2017).

The third paper in this thesis contributes to the contemporary issue of public skepticism towards the news media and science on climate change. It starts with the observation that public proclamations of doubt in the independence of news media are nowadays expressed publicly by politicians. It asks whether these proclamations can motivate consumers to doubt the credibility of news media, and whether such distrust spills over to the science underlying news media articles on climate change. Although the experimental evidence presented in this paper does not allow to make such a conclusion, it nevertheless points towards pathways for further research in such directions.

Methodologically, the three papers presented in this thesis were written to an increasing degree in light of pervasive contemporary discussions about the lack of replicability of experimental findings in (social) psychology and empirical economics, often labeled "replication crisis", or "credibility crisis" (e.g Wagenmakers et al., 2011; Open Science Collaboration, 2015; Ioannidis et al., 2017; Maniadis et al., 2017; McShane et al., 2018). Central causes of the problem are questionable research practices (QRPs) such as p-hacking, optional stopping, selective reporting of variables and experiments, not mentioning experimental conditions, and labeling exploratory as confirmatory hypothesis testing. These practices can spuriously increase the likelihood of finding evidence in support of a hypothesis, albeit at the cost of type I error control (John et al., 2012). These practices can inflate the false-positive rate above the commonly reported threshold of 5 %, consequently leading to non-replicable experimental findings and decreased credibility (Simmons et al., 2011). Also, underpowered studies, primarily as a result of a low number of participants, add to the replication crisis not by causing a high rate of false negatives (the probability of falsely "accepting" the null hypothesis), but of false positives (the probability of falsely rejecting the null hypothesis) (Button et al., 2013). Additionally, a lack of power causes an overestimation of effect sizes (Ioannidis et al., 2017). This results from the file-drawer problem, caused by lacking incentives for researchers to publish results that do not reject the null-hypothesis. Although none of the above procedures are prohibited, they have a negative impact on the positive predictive value, i.e. the probability that a positive research finding reflects a true effect. In the long run, these practices lead to non-reproducible findings, i.e. bad and unreliable science (Button et al., 2013).

Several strategies have been brought forward to counter some of these developments. Some scientists argue for a lower critical p-value in order to label significant findings (Benjamin et al., 2017), while other researchers claim that this misses the problem and other solutions must be found, i.e. transparency and honest reporting (Lakens et al., 2018b), or a complementary continuous- instead of exclusive threshold-interpretation of the p-value (McShane et al., 2018). Contemporary practices, such as preregistration of study design and hypotheses, a priori power analyses, and Bayesian statistics are not direct solutions to these challenges, but they help to alleviate the negative long-term effects and inadequate interpretation of empirical findings. This thesis attempts to abide to a state-of-the-art methodological paradigm facilitating the reproducibility of empirical science insofar as the second and third paper use preregistration and a priori power analyses, while the third paper additionally applies Bayesian statistics to inform the interpretation of null results. Some advantages of preregistering the design, hypotheses, and statistical procedures of an experimental study are that, ultimately, p-hacking is discouraged, ensuring truthfully controlled type I errors (Nosek et al., 2018). Additionally, scientific journals, e.g. Nature Human Behavior, incentivize preregistration by allowing for acceptance in principle, potentially alleviating the file-drawer problem, i.e. not publishing insignificant studies (Nature Human Behaviour, 2017).

Power analyses as such are often used wrongly, i.e after having gathered the data, finding non-significant results, and then using the low observed power as an argument that there was just not enough power to detect the true effect. However, a posteriori power analyses do not provide more information than the p-value, such that the above argument is circular and thus flawed (Hoenig and Heisey, 2001). Consequently, a priori power calculations in the presented papers were used based on prior information about expected effect sizes, in order to recruit the most appropriate number of participants.

The use of Bayes factors in addition to p-values can help to quantify the relative evidence for the null vs. the alternative hypothesis. This is important because conventional null-hypothesis testing does not allow to interpret non-rejection of the null as support for the null (Dienes, 2014). In order to quantify the relative evidence "in favor" of the null hypothesis, Bayes factors can be helpful. The third paper uses this approach to inform its conclusions.

The following four chapters describe the experiments in which the outlined methodologies have been applied to, as well as a conclusion. The thesis is structured as follows: Chapter 2 will present experimental evidence concerning the effect of transparency on the effectiveness of a pro-environmental, and more generally pro-social, default. Chapter 3 presents field-experimental evidence on the performance of a pro-environmental default in relation to conventional instruments, as well as their interaction with intrinsic motivation, and with source information. Chapter 4 presents experimental evidence on the causal role of distrust-statements on trust in a news- and scientific article on climate change. Chapter 5 concludes.

2 Can nudges be transparent and yet effective?

Authors: Hendrik Bruns, Elena Kantorowicz-Reznichenko, Katharina Klement, Marijane Luistro Jonsson, Bilel Rahali
Published April, 2018 in the Journal of Economic Psychology, 65, 41-59.
The data and scripts in support of the findings are openly available at OSF under DOI 10.17605/OSF.IO/569YZ.

Abstract: Nudges receive growing attention as an effective concept to alter people's decisions without significantly changing economic incentives or limiting options. However, being often very subtle and covert, nudges are also criticized as unethical. By not being transparent about the intention to influence individual choice they might be perceived as limiting freedom of autonomous actions and decisions. So far, empirical research on this issue is scarce. In this study, we investigate whether nudges can be made transparent without limiting their effectiveness. For this purpose, we conduct a laboratory experiment where we nudge contributions to carbon emission reduction by introducing a default value. We test how different types of transparency (i.e. knowledge of the potential influence of the default, its purpose, or both) influence the effect of the default. Our findings demonstrate that the default increases contributions, and information on the potential influence, its purpose, or both combined do not significantly influence the default effect. Furthermore, we do not find evidence that psychological reactance interacts with the influence of transparency. Findings support the policy-relevant claim that nudges (in the form of defaults) can be transparent and yet effective.

2.1 Introduction

Nudges, a concept coined by Thaler and Sunstein (2008), describe a diverse set of instruments that utilize behavioral insights in order to affect individual behavior, without limiting options or significantly changing economic incentives. They have become an alternative to economic interventions. While nudges affect behavior by changing the context, thus primarily focusing on automatic decision processes, incentives can be seen to change cognition instead, thus focusing on conscious decision making (Dolan et al., 2012). The recent success of this approach is as a direct consequence of conceiving individual behavior as bounded, instead of perfectly rational and selfish (Bolton and Ockenfels, 2012). Nudges are evolving into a popular form of soft regulation in various fields such as health, finance, and environmental protection (Sunstein, 2014a; Alemanno and Sibony, 2015; World Bank, 2015; Lourenco et al., 2016). Despite its growing popularity, the use of behavioral insights in policy-making is subject to criticism (e.g. Hausman and Welch, 2010; Rebonato, 2014). One remarkable and often criticized aspect of nudges is that they often influence individual behavior without being noticed by the affected subject (Dhingra et al., 2012; Hansen and Jespersen, 2013; Sunstein, 2016). This raises the concern that nudges covertly violate individual autonomy and are therefore unethical (Bovens, 2009; House of Lords Report, 2011). Such regulation thus lacks the transparency that characterizes other regulatory instruments. For instance, when the government imposes a tax to reduce consumption of a product (e.g. cigarettes, or carbon dioxide), people are aware of this tax and can compel the government to justify it (Sunstein, 2014b). On the other hand, when the government sets an opt-out system instead of an opt-in system to promote certain behavior (e.g. organ donation) it exploits several psychological biases, often without people's awareness (Hansen and Jespersen, 2013). Felsen et al. (2013) demonstrate in a vignette study that a significant proportion of individuals have reservations towards nudges they perceive as covert. Additionally, another recent research stream provides evidence of the intrinsic value of decision rights and autonomy (Fehr et al., 2013; Bartling et al., 2014; Owens et al., 2014). To address this criticism, we investigate whether nudges can be made transparent without reducing their effectiveness. In this context, we take into account that the covert nature of nudges is often said to be essential for their effectiveness (Bovens, 2009; House of Lords Report, 2011). Also, we acknowledge that telling people that the nudge is used to influence their decision potentially evokes a perceived threat to their freedom, leading them to experience psychological reactance. The latter can be defined as "the motivational state that is hypothesized to occur when a freedom is eliminated or threatened with elimination" (Brehm and Brehm, 2013, p. 37). This could not only inhibit the effect of the nudge but could even lead to the opposite effect than the one intended. We presume that experiencing reactance is mitigated when information on its purpose substitutes or complements the nudge. According to salience theory (Bordalo et al., 2012), providing the purpose increases the degree to which the ultimate goal of the nudge, relative to its means of behavioral influence, is

taken into account during the decision process. This hypothetically reduces the propensity to elicit a state of psychological reactance. Therefore, this phenomenon is important when investigating the influence of different types of transparency on the effectiveness of nudges. We report evidence from a laboratory experiment where subjects can contribute to real climate protection. The nudge is a default value that intends to increase contributions. Such a default in a public goods context, unlike nudges aiming to improve individual outcomes, attempts to increase positive external effects that only benefit the individual in the aggregate, but affords them to forfeit immediate personal economic gains.[1] Thus, this context is more likely to produce a state of psychological reactance and is thus suitable for testing it.

In general, there are different mechanisms through which a default potentially influences behavior, e.g. as a reference value and anchor (for construction of preferences), through provision of social norms or information, or through inertia (by imposing pecuniary or cognitive costs on deviating from the default). Sunstein and Reisch (2016) provide a review on default-mechanisms. Note that Cappelletti et al. (2014) provide evidence from a public good game that defaults do not work as recommendations, i.e. as information provision in such a context. We expect the default value to increase contributions through two possible ways. First, it can increase the fraction of people picking the default value. Second, it can induce people to increase their contribution towards this value. We discuss our possible mechanisms in the second section and relate them to our findings in the last section.

The type of transparency that accompanies the default varies across treatments and consists of either informing decision makers about its potential behavioral influence and/or informing them about its purpose to increase contributions to climate protection. After the experiment, we assess two different measures of psychological reactance. Thus, we test whether the influence of transparency is limited to a sub-group of participants distinct in their proneness to show psychological reactance (trait reactance). Additionally, we test whether transparency influences the perception of a nudge as a threat for freedom of choice, and whether it functions as a source of anger (state reactance).

Recent findings from Arad and Rubinstein (2017) illustrate why our investigation of transparency and psychological reactance in the context of nudges is important. Their findings suggest that some subjects may consciously act contrary to the encouraged action, presumably in order to protest against the intervention of the government. The authors argue that full transparency of nudges, thus, may even lead to the opposite outcome than the one intended (as opposed to simply eliminating the effectiveness of a nudge). Some people behave in a completely different way simply out of protest against being manipulated. Contrary to this argument, findings by Sunstein (2016) from a nationally representative survey in the USA show that there is widespread support for nudges, and that transparency concerning the nudge

[1] Hagman et al. (2015) divide nudges into pro-self and pro-social. While the former nudge people towards making better decisions for themselves, the latter nudge people towards behavior that benefits society.

will not diminish its effectiveness. Reisch and Sunstein (2016) show that there is also a general support of nudges in six European countries.

To the best of our knowledge, there are only three empirical studies directly relevant to our research question. Loewenstein et al. (2015), in a laboratory experiment, find no evidence that informing subjects that they were presented with a pro-self default option influences their effectiveness. Similarly, Kroese et al. (2016), in a field experiment, find no evidence that making subjects aware of the purpose behind a pro-self default has any effect. Steffel et al. (2016), in several hypothetical and marginally incentivized consumer-related experiments, find no evidence that stressing the potential behavioral influence of a pro-self, as well as a pro-social default impacts their effectiveness, although it affects perception by the consumer.

While existing evidence unanimously suggests the impact of transparency on effectiveness of nudges is absent, our research augments the extant literature in various ways. First, subjects in our experiment face a trade-off between real monetary payoffs and real contributions to a (global) public good. By contrast, two of the previous studies employed relatively abstract and stylized environments and did not demand subjects to make (substantial) financial tradeoffs. Although Kroese et al. (2016) investigate behavior in the field, they do neither study pro-social nudges, nor do they incorporate both types of transparency. Second, we investigate the distinct, as well as combined effect of two types of transparency on the default effect. Previous research focused exclusively on either of these two categories. However, there are reasons to expect that informing decision makers about the potential behavioral influence of a nudge has different consequences than informing them about its purpose. Third, we enrich our analysis with the concept of psychological reactance, allowing for a deeper understanding of potential channels through which transparency influences default effects. Recent research on nudges, although focusing conceptually on the role of reactance (Arad and Rubinstein, 2017; Hedlin and Sunstein, 2016), did not investigate its interaction with transparency.

Consequently, we contribute to the topic of transparency of nudges in various ways. First, we enable a more nuanced view by investigating two types of transparency, thus contributing to a better understanding on how transparency works and whether policy-makers can make nudges more transparent without diminishing effectiveness. Second, our experimental setup, albeit controlled, sets up a realistic context, enabling us to make more valid inferences about the impact of transparency on nudges in "the real world". Third, we widen the discussion on transparency by investigating its connection to the concept of psychological reactance.

To preview our results, defaulted contributions are significantly higher than in the control group, even when accompanied by either type of transparency, including both types. In addition, contributions in the treatment groups (with or without transparency) do not significantly differ from each other. Thus, we replicate the lack of an effect of transparency, indicated by evidence from the studies outlined above. Finally, we neither find evidence that trait reactance interacts with transparency, nor that transparency changes the

perception of nudges as freedom threatening or sources of anger. Therefore, our findings advocate that nudges (in the form of defaults) can be transparent and effective.

The remainder of the article is structured as follows. In Section 2.2 we discuss psychological reactance as a conceptual background to covert nudges, followed by derivation of behavioral predictions. We lay out the experimental design in Section 2.3. In Section 2.4 we present and analyze the results. Section 2.5 concludes.

2.2 Conceptual framework and behavioral predictions

Since Brehm (1966) introduced the theory of psychological reactance, many studies have explored this phenomenon. Social influence attempts (such as nudges) that are detected by an individual may be perceived as a threat to freedom of choice (Brehm, 1966). The elicited state of psychological reactance may result in behavioral and cognitive efforts to reestablish freedom as well as uncomfortable, hostile, aggressive, and angry feelings (Dillard and Shen, 2005). Consequently, people may try to restore their freedom by exhibiting exactly the restricted behavior, thus, in our case, strongly deviating from the default value. In addition, they may devaluate the source of threat (the initiator of the nudge), increase their liking for the restricted freedom, or counter-argue against the imposed option (Brehm, 1966; Dillard and Shen, 2005). People react in such a manner not only to obvious and direct, but also to subtle and subliminal threats (Chartrand et al., 2007).

In order to investigate whether transparency influences the effectiveness of pro-social nudges, specifically defaults, we chose the context of climate protection. With climate change being one of the major challenges faced by society on a global scale today, information-based instruments and nudges are becoming increasingly important to increase individual contributions to climate and environmental protection (Allcott and Mullainathan, 2010; Araña and León, 2013; World Bank, 2015).

One way to contribute to climate protection is to offset (parts of) one's own yearly CO_2 emissions by donating to specific charitable organizations (in the experiment, referred to as "climate protection fund"). These organizations use donations to purchase and delete carbon emission licenses from the European Union Emissions Trading Scheme (EU ETS).[2] Buying carbon licenses is an effective way for individuals to contribute to climate protection, when compared to, e.g. electricity-saving (Perino, 2015). Therefore, individual payment for carbon license retirement is a relevant context in which the influence of transparency on the effectiveness of a pro-social nudge can be investigated.

[2] The EU ETS is a European market that prices carbon emissions and allows regulated industries to trade their emission rights. Buying licenses off the market increases the scarcity of emission rights, resulting in higher prices and thus increasing the incentives for regulated firms to invest in emission-reducing technology.

Based on psychological reactance theory we expect that mentioning the potential influence of a default will evoke the most reactance and thus reduce its effectiveness. In contrast, the sole provision of the purpose, i.e. climate protection, should evoke little reactance since this induces perspective taking. In addition, it renders the positive goal of the contribution more salient. According to salience theory formulated by Bordalo et al. (2012), more salient attributes will be over-weighted in the decision process. Based on this argument, providing the purpose will work as an additional nudge and thus increase the default effect. Finally, accompanying the default with both types of information will be the most transparent form of the nudge. Due to combining the hypothesized "downside" effect of reactance and "upside" effect of the salience of the purpose of the nudge we expect the contribution level to be in between the other treatments. In sum, hypotheses concerning people's contribution decisions in the presence of the default are as follows:

H1: If participants are confronted with a default, contributions will be higher compared to when there is no default.

H2: If participants are informed that the default may have an influence on their decision, contributions will be lower compared to when they are not informed.

H3: If participants are informed of the purpose of the default, contributions will be higher compared to when they are not informed.

H4: If participants are informed of the potential influence of a default and of its purpose, contributions will be higher than with information solely on influence and lower than with information solely on purpose.

Although it is not the purpose of this paper to identify the mechanism underlying the potential default effect, hypothesizing about a transparency-effect relies on certain assumptions regarding this mechanism. Transparency can only exert an effect if subjects are aware of the transparency and consequently of the default. This necessity rules out default effects that rely on unawareness (Madrian and Shea, 2001). If defaults work via costs of opting out (Johnson and Goldstein, 2003), providing a reference point (Samuelson and Zeckhauser, 1988; Dinner et al., 2011) or an anchor (Dhingra et al., 2012), transparency could have an impact.[3] More precisely, information regarding the potential influence of the default then increases the awareness of decision makers to the manipulated structure of the decision. This in turn then may cause reactance. Mentioning the purpose of the default and thus justifying its use has the potential to mitigate reactance. However, note that Wilson et

[3] Note that the potential impact can vary considerably between these mechanisms, and that it can also be close to zero. The point is that here, as opposed to the case of unawareness, transparency could logically influence the default effect.

al. (1996) observe anchoring effects despite forewarning, suggesting an unintentional and subconscious working mechanism that could also apply to defaults working as anchors. If defaults work as an implicit recommendation (McKenzie et al., 2006), a persuasion attempt (Brown and Krishna, 2004), or a coordination device (Cappelletti et al., 2014) it is less clear whether transparency has an effect. Informing decision makers on the potential influence given their interpretation of the default as a recommendation, persuasion attempt, or coordination device would provide no additional information, because decision makers would already be aware of this potential influence. Mentioning the purpose would increase the salience of the climate protection goal, causing a similar effect as when any of the previous mechanisms is at play.

When analyzing findings with respect to psychological reactance, we hypothesize that trait reactance interacts with the type of transparency accompanying the default value. Specifically, we expect that:

H5: If participants are informed that the default may have an influence on their decision, the default effect for participants with higher trait reactance will be lower than for participants with lower trait reactance.

We further hypothesize that the evaluation of a default as freedom-threatening, autonomy-decreasing, manipulative, and pressuring (perceived threat to freedom), as well as its potential to elicit negative emotions (anger) differs with respect to the types of transparency accompanying the default value. Specifically, we expect that:

H6: If participants are informed that the default may have an influence on their decision, experience of state reactance will be higher compared to when they are not informed.

We deduce hypotheses H5 and H6 exclusively with respect to a default accompanied by information on its potential influence, because we expect this type of transparency to increase the salience of the potentially manipulative and autonomy-threatening default-characteristic. For the purpose of the default, the conceptual link to reactance is less clear. We therefore abstain from formulating specific hypotheses.

2.3 Experimental design

The laboratory experiment consisted of five experimental groups, of which one was the control group.[4] We conducted 11 sessions in the Econ-lab of

[4] Prior to the experiment, pilot sessions were conducted in Germany ($n = 16$), Sweden ($n = 25$), France ($n = 29$) and The Netherlands ($n = 32$). The pilot session in Germany focused on developing the design, which was further improved on and tested among Master students in the Netherlands, Sweden, and Bachelor students in France. The experimental design was not identical in all these pilots. Therefore, findings these sessions are not included in the data analysis.

the Erasmus School of Economics at the Erasmus University Rotterdam, the Netherlands, recruited with ORSEE in June 2016, and additional 15 sessions in July 2017 in the WiSo-lab of the University of Hamburg, Germany, recruited with hroot (Bock et al., 2014). A total of 498 students participated in the experiment using the z-tree software (Fischbacher, 2007). Of these, 53.21% were female, the average age was 23.74 years (median: 23 years), and about half (53.01%) studied economics. More information on the differences between samples from both locations, as well as a disaggregated analysis of effect-differences are provided in Appendix A.2.1.

All participants were randomly assigned to separate computer terminals and were instructed not to communicate. They were given instruction sheets that were read aloud (see Appendix A.1). All participants received an endowment of 10 Euro and were asked to indicate how much (if any) of their endowment they would like to contribute to the "climate protection fund". The remaining amount was their private payoff. After the experiment, they were paid according to their decisions, and contributions were used to retire real carbon licenses from the EU ETS, through donations to "TheCompensators*".[5]

In the control group, participants were presented with a text box where they could enter their contribution in any integer amount between 0 and 10 Euro. Neither a preselected default value for the contribution, nor any additional information were presented. In the other experimental groups, subjects encountered an 8 Euro default contribution in form of a button (see Figures A.1 - A.2 in Appendix A.1). They could either press this button or choose another one that stated "Different amount". In the latter case they were referred to another screen that contained exactly the same information but with the addition of a text box where they could insert any amount between 0 and 10 Euro. In three of four default treatments, the default was complemented by a sentence that induced transparency, respectively on the default's potential influence, its purpose, or both. Table 2.1 shows the exact wording used to provide each type of transparency in the respective treatment group.

The Default+Info transparency message informs subjects about the fact that they may be (subconsciously) affected by the default value. It resembles the wording by Steffel et al. (2016) which they use in order to deploy a default *ethically*. We expect that this wording stimulates the participants defensive systems against the threat to their behavioral autonomy, potentially motivating reactant behavior. The Default+Purpose transparency message informs subjects about the purpose of the default, i.e. increasing contributions to the climate protection fund. The wording implies the existence of a default effect, increases the salience of the purpose and, contrary to Default+Info, causes subjects to focus on the goal instead of the fact that it potentially threatens

[5] "TheCompensators*" is a non-profit association founded in 2006 by researchers from the Potsdam Institute for Climate Impact Research. They offer a way for individuals and firms to compensate for their emissions. With donations, they buy and retire emission rights from the EU ETS. At the end of the experiment, all participants received an email with a confirmation and a certificate of aggregate experimental donations to "TheCompensators*".

their behavioral autonomy. The Default+Info+Purpose combines both messages. Once subjects made their decision, they received information regarding their contribution, their private payoff and the amount of CO_2 that would be retired with the contributed amount.[6]

TABLE 2.1: Experimental design

Experimental group	Default value	Transparency information
Control	No	No information
Default	8 Euro	No information
Default+Info	8 Euro	"Please consider that the preselected default value might have an influence on your decision."
Default+Purpose	8 Euro	"Please consider that the preselected default value is meant to encourage higher contributions for the climate protection fund."
Default+Info+Purpose	8 Euro	"Please consider that the preselected default value might have an influence on your decision. This is meant to encourage higher contributions for the climate protection fund."

Notes: The table reports the experimental group, the respective default value presented to participants, as well as the respective transparency information as it was shown to the subjects.

After making their decision, participants answered a questionnaire measuring, among others, their attributed importance to climate protection, and their belief in the effectiveness of retiring emission rights as a measure to protect the climate. In order to find out whether reactions to the different types of transparency can be explained by psychological reactance, we have two approaches. First, we assess participants' perception of the default value as freedom threatening, autonomy-decreasing, manipulative, and pressuring, as well as its tendency to evoke negative emotional reactions, such as irritation, anger, annoyance, and aggravation. We refer to this as state reactance (Dillard and Shen, 2005). Second, we measure subjects' proneness to psychological reactance, referred to as trait reactance, with Hong's Psychological Reactance Scale (Hong and Faedda, 1996). Both measures were assessed after subjects made their decision of how much to contribute.[7] Relevant questions are in Appendix A.3.

[6] At that time, "TheCompensators*" offered to retire licenses at a price of 5.53 Euro. Note that this price can be different from the actual spot-price at the time we conducted the experiment, since "TheCompensators*" buy batches of licenses at a specific price and then retire them based on the donations they receive, irrespective of price-changes that appear in the meantime.

[7] We assume that measuring reactance items before treatments would have introduced an "additional nudge" with a potential influence on contributions. Kruskal-Wallis tests and

After conducting the sessions in Rotterdam, we calculated observed power for the most important tests. For H1, simulated post-hoc observed power analyses produced power coefficients of 0.72, 0.26, 0.51, and 0.46, respectively for Control vs. Default, Control vs. Default+Information, Control vs. Default+Purpose, and Control vs. Default+Info+Purpose. Concerning Findings 2-4, post-hoc observed power analyses for the estimates in model (1) produced power coefficients of 0.22, 0.87, 0.95, respectively for Default vs. Default+Information, Default vs. Default+Purpose, and Default+Info+Purpose vs. Default+Information vs. Default+Purpose. In order to further substantiate Finding 2, we conducted additional sessions for the Control group, Default, and Default+Information groups. The number of additional observations based on an a priori power analysis. The simulation suggested that pooling data from all sessions allowed to detect a true difference of roughly 1.15 EUR (Cohen's $d = 0.37$) in mean contributions between the Default and Default+Information group 78.81 % of the time.

2.4 Results

We present and discuss findings in the following way: First, we demonstrate main results regarding the effectiveness of defaults and their interrelation with transparency. Second, we analyze the measures used to investigate the relevance of psychological reactance to transparency of defaults.

2.4.1 Default effects

Overall, 498 subjects contributed 1,385.5 Euro to retire carbon licenses, resulting in 2.78 Euro per subject. Of all participants, 68.27% contributed a positive amount, and 9.44% opted for the default value. Table 2.2 presents summary statistics of the variables divided by experimental groups. Figure 2.1 presents the respective mean contributions.

A Mann-Whitney test of H1 rejects the null hypothesis of equal contributions between Control vs. Default ($W = 5486, p = 0.001$), Control vs. Default+Info ($W = 4974, p < 0.001$), Control vs. Default+Purpose ($W = 1275, p = 0.032$), and Control vs. Default+Info+Purpose ($W = 1376.5, p = 0.046$). Overall, we find evidence for a default- and pull-effect.

To check robustness of the default effect we focus on contributions as an outcome variable in Tobit regression. The Tobit model accounts for left-censored contributions and allows testing effects on the latent, unobserved contribution variable. This means we assume that at least some subjects would choose to take from instead of contribute to the public good. Thus, we interpret the dependent variable as desired contributions, and indeed even damages, to climate protection. This assumption is common in dictator-games and empirically valid (Engel, 2011).

Steel-Dwass-Critchlow-Fligner multiple comparison tests do not show any significant difference between treatments for all state and trait reactance items. This suggests there is no significant effect of treatments. However, we cannot completely exclude a potential common impact of all treatments on reactance.

TABLE 2.2: Descriptive statistics of all outcome variables to assess the default effect

	Contribution		Contributed	Picked default	n
Group	Mean	SD	Mean	Mean	
Control	1.82	2.66	51.76	0	85
Default	2.95	2.98	70.76	12.28	171
Default+Info	3.04	2.98	74.07	8.02	162
Default+Purpose	2.92	3.19	71.79	15.38	39
Default+Info+Purpose	2.85	2.95	65.85	17.07	41

Notes: The table reports summary statistics (means and standard deviations) of different outcome variables, as well as the number of subjects per experimental group. Outcome variables are: contributions to the climate protection fund, the percentage of subjects contributing a positive amount, as well as the percentage of subjects contributing the default value.

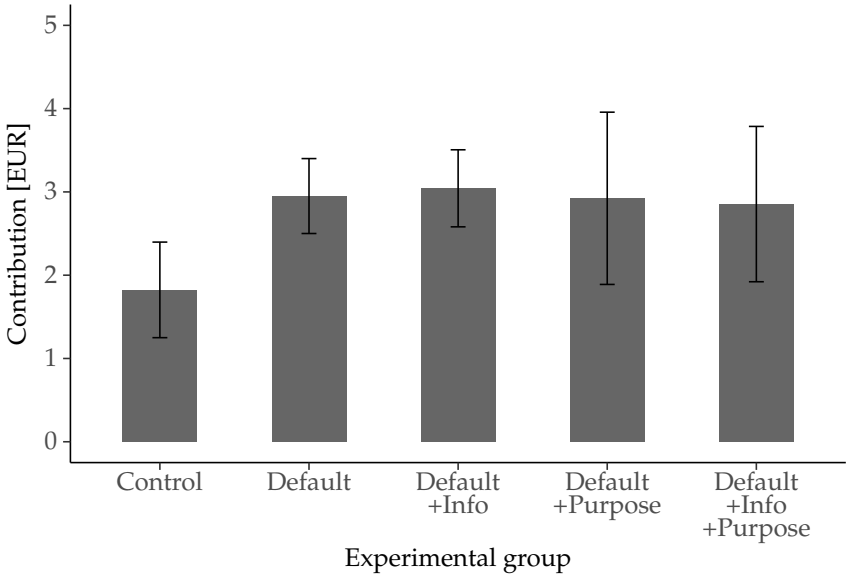
We begin with a restricted model limited to the treatment variable, then add a dummy variable indicating that subjects perceive climate protection to be (very) important, and proceed to add other relevant covariates shown in Table 2.3. The reason we add importance to protect the climate separately is that a Chi^2-Test rejects the hypothesis that subjects are equally distributed among the treatment groups with respect to this variable ($\chi^2(4) = 34.37, p < 0.001$).

By controlling for this variable, we ensure that estimates of treatment effects are not conditionally biased. Because the questionnaire is taken by subjects after being exposed to treatments, there is a risk of the respective manipulations being the reason for the differences in importance-ratings. Regarding Tobit models in Table 2.4, un-restricted model (3) includes all covariates, i.e. rating of the importance of climate protection, gender, age, no previous experience with experiments, judgment of buying emission licenses from the EU ETS as an ineffective tool for climate protection, and a location dummy.

Model (1) predicts that a mere default, a default plus info, and a default plus its purpose lead to higher average contributions compared to no default. The effect of Default+Info+Purpose is marginally significant. When controlling for subjects' perception of the importance of climate protection in model (2), coefficients change. This results in significance for Default+Info+Purpose. *Importance of CP* positively predicts the latent contribution variable. A likelihood-ratio test suggests that model (2) fits the data significantly better than model (1) ($\chi^2(1) = 33.09, p < 0.001$). Controlling for additional covariates increases precision of the estimated average treatment effects. A likelihood-ratio test suggests that un-restricted model (3) fits the data significantly better than restricted model (2) ($\chi^2(5) = 66.40, pp < 0.001$).

F1: There is a default effect on contributions for a default, a default plus information, a default with added purpose, as well as for a default with both types of transparency.

FIGURE 2.1: Mean contributions per experimental group



Notes: The figure shows mean contribution levels in the experimental groups. Error bars represent 95% confidence intervals.

2.4.2 Influence of transparency on default effectiveness

A Kruskal-Wallis test for equal contribution distributions in the treatment groups is not significant ($H(3) = 0.484, p = 0.922$). So are respective pairwise comparisons with Dunn's test (not reported). Consequently, there is no evidence for either of H2, H3, and H4.

As above, we augment our analysis by focusing on contributions in stepwise Tobit-regression (Table 2.4). In un-restricted model (3), an omnibus Wald-test for equality of parameter estimates for Default, Default+Info, Default+Purpose, and Default+Info+Purpose does not lead us to reject the null hypothesis ($F(3, 488) = 0.49, p = 0.692$). The same holds for the restricted models. There is no evidence of unequal contributions in the treatment groups. Consequently, there is no evidence that transparency significantly reduces contributions.[8]

F2: Informing participants that the default may have an influence on their decision does not significantly decrease contributions compared to when they are not informed.

F3: Informing participants about the default's purpose does not significantly increase contributions compared to when they are not informed.

F4: Informing participants that the default may have an influence on their decision, as well as of the default's purpose does not decrease or increase

[8] Estimated treatment-effects of un-restricted regression models are plotted in Appendix A.2 (Figures A.4, A.5, and A.6).

TABLE 2.3: Descriptive statistics of covariates

	Age		Gender (Male)	Importance of CP	No exp. Experience	EU ETS not effective
Experimental group	Mean	SD	Mean	Mean	Mean	Mean
Control	23.75	4.94	48.24	76.47	23.53	60
Default	24.16	4.29	43.27	82.46	29.82	60.23
Default+Info	23.92	4.53	45.06	88.27	25.93	56.79
Default+Purpose	22.28	4.65	53.85	51.28	20.51	64.1
Default+Info+Purpose	22.68	3.72	58.54	63.41	19.51	58.54

Notes: The table reports summary statistics (means and standard deviations) of different covariates per experimental group. Covariates are: age of participants, percentage of males, percentage of subjects perceiving climate protection as (very) important, percentage of subjects without prior experience with experiments, as well as the percentage of subjects judging license retirement as an ineffective mean for climate protection.

contributions, compared to the other types of transparency (including no transparency at all).

Of the additional covariates, *Gender* and *EU ETS not effective* are significant. Being male, as well as judging the EU ETS as not effective to protect the climate, negatively predict the latent outcome variable. The former finding is consistent with evidence from dictator games (Engel, 2011). Findings on gender differences in public good games are ambiguous, however (Croson and Gneezy, 2009). In the context of real contributions to climate protection, evidence by Diederich and Goeschl (2014), while suggesting that female subjects are less indifferent to climate protection, do not support a higher willingness to pay for emission certificates of women. Findings with respect to age somewhat align with those of Borghans and Golsteyn (2015) who find, in a less restricted sample, that the default effect does vary with age. However, at around 22 years (the mean of our sample) they find a relatively large default effect. This may explain why we find a default effect, but no effect of age.

2.4.3 Psychological reactance and transparency

To test if reactions towards the combination of a default value with different types of transparency can be explained by psychological reactance, we measured subjects' proneness to experience psychological reactance.[9]

[9] To create an index for trait reactance, we constructed dummy variables for each of the 14 items of the scale, which are equal to 1 when the subject responded with "Agree" or "Strongly agree" to the respective question, 0 otherwise. We then added the dummies for each subject to create the index, which ranges from zero to 14. Findings are consistent for trait reactance included as a (un-weighted) factor-based score.

TABLE 2.4: Stepwise Tobit-models with and without interaction term

	(1) Contribution	(2) Contribution	(3) Contribution	(4) Contribution	(5) Contribution	(6) Contribution
Default	1.868**	1.718**	1.659**			
	(0.587)	(0.571)	(0.539)			
Default+Info	2.056***	1.758**	1.670**	0.165	0.0152	0.00216
	(0.586)	(0.577)	(0.538)	(0.438)	(0.429)	(0.410)
Default+Purpose	1.866*	2.612**	2.528**	-0.0343	0.858	0.841
	(0.839)	(0.845)	(0.784)	(0.730)	(0.750)	(0.775)
Default+Info+Purpose	1.628x	1.921*	1.896*	-0.260	0.169	0.174
	(0.829)	(0.779)	(0.779)	(0.726)	(0.670)	(0.756)
Importance of CP		2.806***	2.350***		2.810***	2.353***
		(0.517)	(0.502)		(0.558)	(0.534)
Gender (Male)			-1.045**			-1.065**
			(0.353)			(0.391)
Age			-0.0406			-0.0200
			(0.0403)			(0.0431)
No exp. Experience			-0.577			-0.522
			(0.425)			(0.451)
EU ETS not effective			-2.512***			-2.329***
			(0.347)			(0.368)
Hamburg			-0.0494			-0.102
			(0.453)			(0.504)
React				-0.0897	-0.0977	-0.0783
				(0.106)	(0.102)	(0.0971)
Default+Info × React				-0.108	-0.109	-0.0764
				(0.145)	(0.141)	(0.133)
Default+Purpose × React				0.183	0.208	0.114
				(0.276)	(0.285)	(0.250)
Default+Info+Purpose × React				0.0646	0.0316	-0.0483
				(0.224)	(0.206)	(0.190)
Constant	0.357	-1.824**	1.986x	2.259***	-0.0734	3.072**
	(0.497)	(0.644)	(1.094)	(0.314)	(0.563)	(1.100)
Sigma	3.969***	3.848***	3.591***	3.888***	3.766***	3.550***
	(0.152)	(0.153)	(0.143)	(0.153)	(0.152)	(0.147)
Observations	498	498	498	413	413	413
Log Pseudolikelihood	-1088.416	-1071.872	-1038.671	-929.4	-915.187	-890.107
F	(4, 494)=3.33	(5, 493)=8.64	(10, 488)=13.19	(7, 406)=0.76	(8, 405)=3.98	(13, 400)=7.36
Prob > F	0.010	<0.001	<0.001	0.624	<0.001	<0.001
Pseudo R^2	0.007	0.022	0.052	0.002	0.018	0.044

Notes: The table reports estimates of Tobit models with contributions censored at 0 as the dependent variable, with and without interaction terms. Robust standard errors are in brackets. *Default+Info*, *Default+Purpose*, and *Default+Info+Purpose* denote the respective treatment group, with *Default* as the base category. *React* measures subjects' proneness to experience reactance in a metric scale, and is mean centered. *Def+Inf × React*, *Def+Pur × React*, and *Def+Inf+Pur × React* are interaction terms of the transparency type with proneness to experience reactance. *Importance of CP* is a dummy that takes the value 1 if the subject perceives climate protection as (very) important. *Gender* takes the value 1 if the subject is male. *Age* denotes the age of the subject. *No exp. Experience* is a dummy which takes the value 1 if a subject did not participate in another experiment before. *EU ETS not effective* is a dummy that takes the value 1 when a subject judges license retirement as an ineffective mean for climate protection. *Hamburg* takes the value 1 if the subject is from the Hamburg, as opposed to the Rotterdam sample. Significance levels: ^{x} ($p < 0.10$), * ($p < 0.05$), ** ($p < 0.01$), *** ($p < 0.001$).

Specifically, we test whether subjects' reactions towards different types of transparency accompanying the default differ depending on subjects' trait reactance. Therefore, we run regressions with an interaction term of the treatment variable and the trait reactance index. The latter is centered on the mean, so that treatment-main-effects are meaningful (Table 2.4). Note that this regression excludes observations from the control group. For reasons of brevity, we focus on the main effects of trait reactance, as well as on interaction-effects.

As in previous Tobit models, model (5) fits the data better than model (4) ($\chi^2(1) = 28.42, p < 0.001$), and model (6) fits the data better than model (5) ($\chi^2(4) = 50.11, p < 0.001$). We find no significant main effect of trait reactance, nor do we find that the different types of transparency and the trait reactance index interact significantly for any of the three model-specifications. In other words, there is no evidence that the effect of different types of transparency on average contributions is conditional on subjects' trait reactance.

F5: The influence of information on the default effect does not depend on the level of trait reactance of participants.

In order to test whether reactions to different types of transparency can be explained by psychological reactance, we create an index for each of the two state reactance-categories, i.e. for the perceived threat to freedom and the anger-category.[10]

We model the log odds of subjects being in a higher level of each of both ordinal indexes on all explanatory variables used above (Table 2.5). Note that this regression excludes observations from the control group since subjects in this group were not presented with the default option which they could rate. None of the coefficients modeling treatment effects are significant.[11]

F6: Combining the default with information about its potential behavioral influence does not increase participants' experience of state reactance.

Age negatively predicts experienced anger triggered by the default value. The finding that experiencing negative emotions decreases with age is known in the literature (e.g. Charles et al., 2001). Both approaches that are linking different types of transparency of a default to psychological reactance suggest that subjects neither perceive a default value differently based on the type of transparency accompanying it, nor does their inherent propensity to show psychological reactance change the way they react to these different types of transparency.

2.5 Discussion and conclusion

The experiment advances the discussion of nudges and transparency by providing empirical evidence on the effect of transparency on the performance of a pro-environmental default value. Despite the widespread application of nudges, many researchers and consumers are concerned of the potentially manipulative nature of behavioral interventions. In democratic societies, public authorities are expected to be transparent with regard to their actions and intentions. Therefore, covertly "exploiting" people's psychological biases potentially inhibits perceived legitimacy, and ultimately effectiveness of such policies. The most straightforward solution to this problem is to instruct policy-makers to disclose information regarding the potential influence of

[10] We constructed a dummy-variable, which is equal to 1 when the subject "agreed" or "strongly agreed", resp. replied with "to some extent" or "very" to the respective statements, for each item (see Appendix A.3). Then, we added the respective dummies in each category, to form two indexes, each ranging from zero to four. Findings are consistent for when both dependent variables are included as (un-weighted) factor-based scores in linear OLS-regression.

[11] This finding is consistent with non-parametric tests for differences of individual items of the scales (not reported).

TABLE 2.5: Ordered logistic model of state reactance

	(1) Threat To Freedom	(2) Anger
Default+Info	-0.00294	-0.167
	(0.199)	(0.223)
Default+Purpose	-0.0297	0.0868
	(0.418)	(0.453)
Default+Info+Purpose	-0.0686	-0.560
	(0.330)	(0.470)
Importance of CP	-0.0275	-0.334
	(0.232)	(0.276)
Male	-0.0798	-0.300
	(0.190)	(0.217)
Age	-0.0594**	-0.0832**
	(0.0183)	(0.0268)
Participated	-0.0221	-0.0560
	(0.192)	(0.242)
EU ETS not effective	0.183	0.173
	(0.191)	(0.216)
Hamburg	-0.0120	-0.325
	(0.250)	(0.260)
Cut 1	-3.125***	-2.029**
	(0.528)	(0.683)
Cut 2	-2.270***	-1.126^{x}
	(0.524)	(0.679)
Cut 3	-1.088*	-0.251
	(0.517)	(0.685)
Cut 4	0.346	0.508
	(0.525)	(0.718)
Observations	413	413
Log Pseudolikelihood	-640.583	-443.190
Wald Chi^2(9)	12.96	19.80
Prob $>$ Chi^2	0.165	0.019
Pseudo R^2	0.008	0.024

Notes: The table reports estimates of ordered logit models with ratings of defaults as threatening to freedom, and anger arousing as the respective dependent variable. Robust standard errors are in brackets. *Default+Info*, *Default+Purpose*, and *Default+Info+Purpose* denote the respective treatment group, with *Default* as the base category. *Importance of CP* is a dummy that takes the value 1 if the subject perceives climate protection as (very) important. *Gender* takes the value 1 if the subject is male. *Age* denotes the age of the subject. *No exp. Experience* is a dummy which takes the value 1 if a subject did not participate in another experiment before. *EU ETS not effective* is a dummy that takes the value 1 when a subject judges license retirement as an ineffective mean for climate protection. *Hamburg* takes the value 1 if the subject is from the Hamburg, as opposed to the Rotterdam sample. Significance levels: $^{x}(p < 0.10)$, $^{*}(p < 0.05)$, $^{**}(p < 0.01)$, $^{***}(p < 0.001)$.

the nudge, and its purpose. However, this suggestion raises the concern that nudges will no longer be effective. As expressed by Bovens (2009), nudges "work best in the dark". The results of this study suggest that this concern might be overstated.

The experiment provides evidence that defaults increase contributions to climate protection even when complemented by disclosure regarding the potential influence of the default, its purpose, or both. Furthermore, there is no

evidence that information on the potential behavioral influence and/or purpose of the default triggers psychological reactance. Likewise, there is no evidence that subjects differing in their proneness to experience reactance also differ in how they react towards the default with additional information.

These findings suggest that despite the initial concern over the inhibiting influence of transparency, nudges in the form of defaults can be transparent and at the same time effective. In order to preserve the effect of defaults and increase the legitimacy of behaviorally informed policies, policy makers should be transparent about their motives, as well as the potential behavioral influence of the instrument. The motive and how it is perceived by the decision maker has been found to matter for advice (Kuang et al., 2007).

Our findings replicate and add to previous evidence on the influence of transparency. Loewenstein et al. (2015) and Kroese et al. (2016) reported that pro-self defaults were effective in health contexts even after disclosing information about them. Our study extends this conclusion to pro-social nudges, a type that is widely used in the context of public policy-making. Moreover, we extend findings of Steffel et al. (2016) by examining the influence of transparency in a more realistic setting where participants' decisions have an actual consequence for them, and for the environment. Findings are also useful for the private sector and NGOs aiming to include nudges in their inventory to increase contributions to environmental protection, and possibly other public goods, e.g. charity.

Although several recent studies link nudges to psychological reactance, they do so either indirectly, or they deal with hypothetical and attitudinal, instead of behavioral outcomes (Haggag and Paci, 2014; Arad and Rubinstein, 2017; Loewenstein et al., 2015; Hedlin and Sunstein, 2016). By measuring both state and trait reactance, we enable a more direct way of assessing the interaction of psychological reactance with the influence of transparency on the effectiveness of a default value. To our best knowledge, Goswami and Urminsky (2016) is the only study that assesses the interaction of trait reactance with the size of a default value on behavioral outcomes, i.e. charitable giving. They find no significant interaction effect. On a more general level, our findings, in line with theirs, suggest that psychological reactance plays a minor or no role with respect to behavioral effects of defaults, and, in our case, transparency. In fact, a possible explanation of this might be the relatively high default value, which is 80 % of the experimental endowment. Instead of eliciting psychological reactance, such a high default might lead subjects to ignore it altogether.

Findings suggest that the default value is an effective way of increasing individual voluntary contributions to climate protection. Increased aggregate contributions are consistent with inertia, as well as anchoring. A higher fraction of participants picking the default value instead of specifying another amount in the default, compared to the control group, supports the inertia/ effort reduction explanation. However, deviation costs in the experiment are marginal (the subject had to make two mouse-clicks, as well as to type in the contribution amount, instead of just making one mouse click on the default button), and contributing the default value is also consistent with

an anchoring explanation: Subjects may choose the default value not only because of inertia, but also because they consider this value first and only then employ reasons against it, conditional on what they wanted to contribute initially. This anchoring-explanation is consistent both with picking the default and moving towards the default, whereas inertia is only consistent with the former behavior (Dhingra et al., 2012).

We observe that subjects who contribute a positive amount do contribute more on average, when there is a default value with either type of transparency, but the differences to the control group are not significant. Additionally, we observe an increase of subjects giving a positive amount due to the default, which is consistent with the anchoring explanation. Together, our findings suggest that increased aggregate contributions in the default groups are due to an increase of the fraction of subjects contributing, as well as of an increase of the fraction of subjects choosing the default value, but not because of increased average contributions of subjects that contribute. Inertia, as well as anchoring may therefore both be reasons for why we observe default effects. Intuitively, we would expect anchoring to play a more pronounced role in real world applications of pro-environmental nudges, especially if defaults result in repeated and/or significant financial costs. For someone who highly values environmental- and climate protection, deviating from a default, which may be perceived as conveying information about social norms, can incur non-financial costs, especially if he or she aims to uphold a positive self-image. Maintaining a positive self-image, as well as being consistent with social norms, can be achieved by decreasing (not necessarily closing) the gap between default value and initially intended contribution. Note that our design does not allow to unambiguously identify the underlying mechanisms causing the default effect in the experiment. Anchoring is consistent with the interpretation of the default value as an implicit recommendation, a persuasion attempt, coordination device, or a reference point. If a decision maker regards the default as an implicit recommendation, she may consequently increase/decrease her donation relative to her preferences, after seeing the default. However, we cannot identify whether she interpreted the default as a recommendation.

Furthermore, while being able to differentiate between the effects of different types of transparency is insightful for policy-makers, the difference between the information and purpose treatments is not analytically clear.[12] Communicating the purpose of the default implicitly reveals that the default is expected to have an effect on individual decision making, without spelling it out. Still, we think that the findings concerning this type of transparency are important for practical purposes.

Further research could evaluate the role of trait reactance on how subjects respond to different types of transparency for different types of nudges, i.e. social norms or framing. Additionally, building on the shortcoming of our experimental design, further studies should further investigate the link between transparency and the different underlying working mechanisms of defaults and other types of nudges. Since our experiment has a rather limited

[12] We thank an anonymous reviewer for this remark.

number of subjects, field experiments can establish statistically more powerful findings for interaction effects. Due to a more realistic context, a field experimental approach would also increase external validity. Nevertheless, our experiment is less abstract than a "regular" laboratory experiment due to the fact that contributions have a real effect on climate protection Harrison and List (2004). The current study focuses on one type of nudge, and a specific context. Further research is needed in order to determine the overall influence of transparency on the effectiveness of nudges. Moreover, results might be context-specific, thus requiring further investigation into pro-social nudges. Delving into the welfare implications of transparency can also become a promising research endeavor (Sunstein, 2015).

Overall, our findings advance the understanding of how nudges in general, and defaults specifically, affect individual behavior with social consequences, and how policy-makers can increase their transparency without limiting their effectiveness.

3 Point at, nudge, or push private provisions to a public good? Field experimental evidence for experts, politicians, and nobodies

Authors: Hendrik Bruns, Grischa Perino
The data and scripts in support of the findings are openly available at OSF under DOI 10.17605/OSF.IO/6BDNK.

Abstract: This paper investigates differences between a default, a recommendation, and a mandatory minimum contribution on private provision of a large scale public good (climate protection). Information on the regulator (neutral experimenter, expert or politician), its interaction with the intervention type, as well as with pre-intervention intrinsic motivation on voluntary contributions is analyzed. Data are from an online framed field experiment with a sample representative of the German internet using population. Main insights are: neither a recommendation nor a default close to the pre-intervention average change contributions; identifying the regulator reduces contributions when accompanying the recommendation, but not the default; contributions above the pre-intervention average are reduced by the default but increased by the mandatory minimum relative to the control; only the default negatively interacts with high intrinsic motivation; and regulator attributes neither interact with intrinsic motivation, nor with the type of intervention. The study contributes to the discussion of nudges as public policy instruments by comparing them to alternative interventions, i.e. pointing at or pushing contributions, and by shedding light on making the source of the intervention transparent.

3.1 Introduction

Nudges are interventions that alter behavior without using financial incentives or limiting choice options (Thaler and Sunstein, 2008). At the same time, they go beyond the mere provision of information, aiming not at attitudinal, but rather behavioral outcomes as the target of influence. Governments and private actors increasingly use them to substitute or complement information provision campaigns, taxes and subsidies, or mandates (Lourenco et al., 2016). Particularly, nudges are means to motivate voluntary private contributions to public goods, like environmental- and climate protection (World Bank, 2015; Bühren and Daskalakis, 2015), or charity (Altmann et al., 2014; Goswami and Urminsky, 2016). Nudges allegedly alleviate the problem that soft interventions, such as information provision, do not reliably translate attitudinal changes to behavior (Croson and Marks, 2001; Kollmuss and Agyeman, 2002). Nudges, per definition, attempt to achieve this by predictably changing behavior instead of attitudes, via deliberate and autonomy-preserving changes in the decision context. Additionally, they owe their prominence relative to traditional economic incentives, including restrictions, to the latter's potential to negatively interact with decision makers' intrinsic motivation (Cardenas et al., 2000; Falk and Kosfeld, 2006; Meier, 2007; Reeson and Tisdell, 2008; Goeschl and Perino, 2012; Perino et al., 2014). As of yet, although nudges are not without criticism concerning their effectiveness (Jachimowicz et al., 2017), there is limited evidence that their effectiveness depends on the intrinsic motivation of decision makers.

Yet, recent research highlights the intrinsic importance of behavioral autonomy, e.g. of preserving decision rights for decision makers (Fehr et al., 2013; Bartling et al., 2014; Owens et al., 2014). Restricting choice autonomy and decision rights has been established by Falk and Kosfeld (2006) as a cause for motivation crowding out, defined as an intervention's negative effect on intrinsic motivation (Bowles, 2008). Our paper is the first to answer how nudges, more specifically defaults, differ from recommendations, and restrictions of choice options in their effect on and interaction with intrinsic motivation to contribute to a public good.

A second concern with nudges and interventions more generally is the influence of regulator-attributes, i.e. their characteristics perceived by the decision maker, on the effects of these instruments. For example, regulators' party affiliation has been shown to affect how decision makers respond to nudges (Tannenbaum et al., 2017), perceived motives of the policy maker have been shown to influence the effects of defaults and recommendations (Brown and Krishna, 2004; Kuang et al., 2007), informational asymmetries between default setter and decision maker influence default effects (Altmann et al., 2015), and financial incentives backfire when perceived as governmental instead of neutral interventions (Perino et al., 2014). Additionally, findings by Falk and Kosfeld (2006) show that motivation crowding-out is affected by whether or not the imposition of control results from a regulators' conscious decision, or from an exogenous experimenter. Consequently, we investigate the effect of revealing the identity of the regulator on and interaction with

intrinsic motivation to contribute to a public good.

We present data from an online framed field experiment with a sample representative of the German internet using population, in which participants act as decision makers facing a combination of intervention and real regulator-attribute, also measuring a proxy for intrinsic motivation prior to treatment allocation.

Major findings are: (1) Neither recommendation nor default close to the pre-intervention average affect average contribution changes relative to a control without an intervention. (2) Revealing the identity of the regulator reduces average contributions for the recommendation, but not the default. (3) Contributions above the pre-intervention average are reduced by the default but increased by the mandatory minimum relative to the control. (4) The default negatively interacts with *high* intrinsic motivation. This is neither the case for a mandatory minimum contribution, nor a recommendation. (5) Regulator attributes do not interact with intrinsic motivation for either type of intervention. (6) The relative effect of interventions does not depend on the type of regulator responsible for its implementation.

While the main insight is the comparative crowding out of intrinsic motivation due to a default, the absence of a statistically significant recommendation-, and default-effect on aggregate contributions, taking into account intrinsic motivation, is equally important in light of the recent praising of nudges as effective and efficient policy alternatives (e.g. Benartzi et al., 2017). Additionally, while the positive effect of mandatory minimum contributions on average contributions is by design, contributions above the contribution threshold are in contrast to related findings by Falk and Kosfeld (2006) still increased relative to the control, rendering this intervention the most effective in raising contribution levels. Hence, the study contributes to the literature on the relation between regulator-information and intervention effect, as well as to the literature investigating the potential of behavioral interventions to induce unintended effects.

To the best of our knowledge, we are the first to provide direct and controlled experimental evidence on the relative performance of these three types of interventions, the role of regulator-attributes, as well as their interaction with intrinsic motivation to make real contributions to a large public good. Findings therefore help to better understand the distinction between nudges and other ways to affect behavior (Hansen, 2016). Outcomes are insightful for regulators and policy makers insofar as they help predicting who should prefer to use which intervention, i.e. a pointer, a nudge, or a push, in order to increase private contributions to a public good.

The remainder of this article is structured as follows: section 3.2 outlines the experiment, outcome measures, behavioral predictions, participants, and statistical analyses used. Section 3.3 presents the data analysis and hypotheses tests. Section 3.4 discusses important caveats of the design, while section 3.5 concludes.

3.2 Experiment

The experiment was conducted online with a sample drawn from the panel of a professional survey and market research company in Germany. The basic choices resemble a modified dictator game, where the recipient of monetary contributions is climate protection representing a large scale public good as in e.g. Diederich and Goeschl (2017). Data collection is divided in two stages and three elements: The first element corresponds to Stage 1 and consists of a questionnaire conducted about three weeks prior to Stage 2. The latter had two elements: the incentivized experiment and an exit questionnaire. Importantly, Stages 1 and 2 were set up in a way that participants could not link the two stages. As they are registered panel members of a professional survey company, being contacted to participate in a survey in itself is not sufficient to create a link between the two stages.

Here, the description of the experiment is limited to Stage 2. As part of the experiment, subjects made two contribution decisions, of which the second was modified according to the treatment they were randomly assigned to. According to a Random Lottery Design (RLD), only one of the two decisions was realized. After the experiment, we donated one randomly chosen contribution of each participant to the non-governmental organization "TheCompensators*"[1], which buys and retires emission rights from the European Union Emissions Trading System (EU ETS). The experiment was implemented online with Limesurvey, a tool to design and conduct online surveys. Participation was possible via personal computers and smartphones from any place with internet access. The experimental design was registered and made available online at the American Economic Association's registry for randomized controlled trials (AEARCTR-0001661).[2]

3.2.1 Experimental procedure and treatments

Subjects stated their age, gender, and education, before reading the instructions. After a comprehension test question, they were endowed with 100 Credits ($\sim$ 5 EUR or $\sim$ \$6) and made a baseline contribution decision. Subjects where then randomly allocated to treatments, where they encountered either a recommendation, a default, or a mandated minimum amount, each of 35 Credits to contribute to climate protection. The respective intervention was combined with either no information on the source responsible for the implementation of the intervention, with information labeling the source as an expert on environmental science and climate change policy, or with information characterizing her as a politician. Note that the information provided on the regulator is factually true and represent different stages of her career. Furthermore, during the experimental design phase she specified the value

[1] "TheCompensators*" is a German non-profit association founded in 2006 by researchers from the Potsdam Institute for Climate Impact Research. It offers a way for individuals and companies to compensate for their emissions and thus contribute to climate protection.

[2] In April 2016, we conducted a laboratory pilot experiment in the experimental research lab at the University of Hamburg. Findings from these sessions were used to develop and improve the experiment.

at which the interventions were targeted, i.e. 35 credits out of 100, which she could choose from a range of contribution levels pre-specified by us. That the targeted level turned out to be close to the pre-intervention average of contributions was a coincidence as similar studies report lower mean and median contributions (Diederich and Goeschl, 2014; Diederich and Goeschl, 2017; Bruns et al., 2018).

Table 3.1 shows the labels of the respective experimental groups, i.e. the combinations of factor levels, from the fractional factorial design. Subjects in the control group made the same decision twice without any intervention. Before making the baseline contribution, subjects where informed that the payout-relevant contribution decision would be randomly selected at the end. The decision space for contribution c of all subjects was identical in the baseline decision, i.e. $c \in [0, 100]$. In the second, treated decision, the decision space differed only for the case of the minimum mandatory contribution, where subjects could specify a contribution $\underline{c} \in [35, 100]$.

TABLE 3.1: Experimental design

	Intervention type			
Source type	No Intervention	Recommendation	Default	Restriction
No regulator	Control	Rec-No	Def-No	Res-No
Expert regulator	n/a	Rec-Exp	Def-Exp	Res-Exp
Political regulator	n/a	Rec-Pol	Def-Pol	Res-Pol

Notes: Shows the labels for treatment groups.

In treatments with information on the regulator's identity, this was presented to subjects after they had made their baseline choice in the following sequence: first a screen with a picture of the person and a short paragraph about her background (expert or politician), followed by the intervention and contribution screen that linked the intervention to the person and also gave the option to review the picture and career info. Participants were told that the person was real, that she was involved in the design of the experiment, and that her name was revealed at the end of the experiment. The detailed paragraphs about the expert, resp. political regulator were, respectively:

Expert: *"This person received a PhD on the topic of 'Energy-efficiency policy as a contribution to climate protection' from the University of Lüneburg, graduating as Dr. phil. Before that, she received a master's degree in environmental sciences from the same University. For several years, she worked as a research associate in this field."*

Politician: *"This person is a politician. She is a member of the German Federal Parliament and an ordinary, resp. surrogate member in numerous committees and commissions. In the past, she was executive director of an administrative district for her party. Additionally, she worked as a delegate and councilor in a district- and city*

council of a German city with around 70.000 inhabitants."

Subjects in non-regulator-information treatments skipped the corresponding screens and directly proceeded to the contribution decision that also contained the treatment intervention. Subjects in the control group saw an interface identical to the first decision. Subjects in the recommendation group were presented an interface consisting of a sentence that the respective regulator (if any) recommended them to contribute 35 Credits. In the default group, they were confronted with two radio buttons, of which the option to contribute 35 Credits was pre-chosen. Subjects were free to choose the second button which allowed them to specify another amount. This set-up was accompanied by a sentence that the respective default was implemented by the person previously presented to them (if any). When facing a restriction on possible contributions, they saw a text-box identical to the control- and recommendation groups. However, they could only specify an amount between and including 35 and 100 Credits. Here again, the intervention was linked to the respective regulator (source), if any was provided. In each of the cases where no source information was provided, the interventions were introduced neutrally by stating "35 Credits were set as a default/mandatory minimum contribution", resp. "It was recommended to contribute 35 Credits". After the contribution decision, subjects were informed about the consequences of their realized decision. Subjects then answered a short post-experimental questionnaire. Instructions and screen-shots are in Appendix B.1.

3.2.2 Outcome measures

The main outcome of interest as specified in the pre-analysis plan is individual contribution to climate protection, measured in Credits. The within-subject design allows to construct a contribution change variable by subtracting the first-round contribution from the (treated) second-round contribution. We include the first-round contribution as a measure of the latent intrinsic motivation to explain the deviation of final (treated) contributions from intrinsic motivation (Allison, 1990).

3.2.3 Behavioral predictions

Based on previous research we expect that the change in subjects' contributions between the first and the second decision depend on the intervention they encounter in between and their ex-ante level of intrinsic motivation, i.e. the contribution level in decision 1. For example, Bruns et al. (2018) show that a plain default, as well as a default complemented by information on its purpose and likely influence, increases individual contributions to climate protection. For recommendations, i.e. advice or "cheap talk", Croson and Marks (2001) provide evidence for its effectiveness in threshold public goods game. Cappelletti et al. (2014) show in a public goods game that preference

for a recommended contribution increases when it is set as the default, indicating that a default effect is stronger than plain advice.

Only a restriction on choice options objectively limits choices and should therefore lead to control aversion. In general, research on motivation crowding is concerned with the effects of economic incentives and restrictions of choice on intrinsic motivation, e.g. to contribute to a public good.[3] Varying the default level, Goswami and Urminsky (2016) report evidence on a "scale-back" and a "lower-bar" effect for donation-defaults, describing the situation when low defaults decrease average contributions, and increase the number of subjects choosing the default value, respectively. Contrary to our experiment, they do not observe intrinsic motivation, and hence cannot observe the difference between default value and intrinsic motivation for each individual. To the best of our knowledge, this procedure is new for nudges and enables us to establish a link to motivation crowding.

In Falk and Kosfeld (2006) crowding out by a restriction of the choice set occurs when implemented by a principal who personally benefits from higher contribution levels but not when implemented by the experimenter. The latter is in line with results by Goeschl and Perino (2012) in the context of a large scale public good. Cardenas et al. (2000) provide field-experimental evidence that exerting control by imposing an upper-limit on resource extraction in an environmental dilemma-situation increases self-interested behavior, which they interpret as crowding out of other-regarding preferences. Reeson and Tisdell (2008) describe similar behavioral patterns in a laboratory public goods game with mandatory minimum contributions. They observe crowding out of voluntary contributions after discontinuing the regulation. Although Fitzsimons and Lehmann (2004) provide evidence that psychological reactance, a cognitive or behavioral reaction to perceived threat to freedom, potentially eliciting behavior similar to motivation crowding-out, can occur as a reaction to recommendations, these interventions, as well as defaults, per definition, do not interfere with choice options and are therefore hypothesized to not interact with intrinsic motivation. Based on this evidence, we state the following hypotheses:

H1a: Recommendation and default induce an increase in contributions from decision 1 to 2 relative to the control group.

H1b: Contributions above 35 are expected to decrease relative to the control group in the mandatory minimum condition, and more so the higher the initial level of intrinsic motivation (crowding out).

Research also suggests that the source information complementing an intervention affects its performance, and that motivation crowding occurs because decision makers assess the decision situation based on what they know about the institution or person responsible for implementing the intervention (Bénabou and Tirole, 2003). For example, Brown and Krishna (2004)

[3] For a general review of research on motivation crowding see Bowles and Polanía-Reyes (2012). Rode et al. (2015) provide a review in the context of conservation policy.

provide experimental evidence that the default effect decreases when the default setter conveys the impression that he or she follows his or her own profit-maximization agenda at the expense of the decision maker. Results indicate that decision makers treat defaults as informative about default setters' intentions. Altmann et al. (2015) enrich this by providing evidence from a laboratory experiment that the effect of a default on behavior depends on how information and interest with respect to the choice outcomes differ between decision maker and default setter. Their underlying model, which is consolidated by the data, suggests that, if interests between default setter and decision maker are aligned, defaults are more informative for the latter agent. Additionally, the model suggests that default effects are strongest when decision makers are less informed than default setters. Tannenbaum et al. (2017) show that party affiliation plays a significant role when assessing behavioral policy interventions like nudges: party affiliation of a default setter affects decision makers' ratings of defaults as ethical, coercive, and manipulative. Kuang et al. (2007) report evidence from a pure coordination game that advice is less effective when its source is perceived by decision makers as self-interested. In other words, advice that results in a positive payoff for the adviser and costs for the advisee is less effective than the same advice coming from an external adviser. Some of the experimental findings reported by Falk and Kosfeld (2006), as well as Goeschl and Perino (2012) suggest that the information of the source responsible for restricting choices or imposing a tax plays a role. Evidence presented in Perino et al. (2014) shows that information of the source responsible for a pro-environmental subsidy may change its effectiveness. Consequently, the effectiveness of all types of interventions discussed in this paper, soft-interventions, nudges, and economic instruments, is potentially affected by information of the source complementing it. Also, this information appears to influence factors relevant to motivation crowding. Based on this evidence, we state the following hypotheses:

H2a: Providing source information (differently) affects behavioral responses to the interventions.

H2b: In particular, the provision of information on the regulator responsible for implementing an intervention interacts with intrinsic motivation to contribute to climate protection, and

H2c: The influence of an intervention on contributions to climate protection depends on the type of regulator responsible for its implementation.

3.2.4 Participants

The final data consist of 806 observations with a mean age of 49.90 (SD = 15.66, Median = 51) years, 51.86 % women, with 33.75 % having lower, 33.37

% middle, and 14.76 % higher education, while 17.74 % hold a university degree.[4] Aggregated and disaggregated distributions of these covariates, distributions of answers given to central questions from the questionnaires, as well as a detailed description of data cleaning are in Appendix B.2. The respective questionnaires are in Appendix B.3.

3.2.5 Statistical analyses and power analyses

An a priori power analysis suggested that in order to detect an interaction effect between intervention and provided regulator information with a standardized effect size of Cohen's $f = 0.13$ (medium effect), with a power of 80 %, assuming an alpha level of 0.05, and using a two-way analysis of variance (ANOVA) while excluding the control group (9 groups, $df = 4$) needs $n = 720$. This test is what we powered the study for. A two-tailed Wilcoxon-Mann-Whitney test with the same parameters as above and group sizes of $n = 80$ detects an effect of $d = 0.46$ (medium effect) 80 % of the time. This is adequate, keeping in mind that Jachimowicz et al. (2017) estimate a default effect of $d = 0.59(CI_{95} = [0.45; 0.74])$ in a meta study including 71 default studies. A power analysis for regression model (3) presented below shows that 160 observations (two experimental groups) allow to detect a standardized minimum effect size of $f^2 = 0.05$ (medium effect) for an interaction, i.e. one tested, and 8 overall predictors 80 % of the time (Faul et al., 2009). This calculation does not take into account that change scores increase the statistical power due to a reduction of statistical variance in the dependent variable (Allison, 1990).

All tests are based on a difference-in-difference design. H1a-H2a are tested with non-parametric Wilcoxon Mann-Whitney tests. H1b and H2b/c are tested with the following OLS regression models:

$$(c_{2i} - c_{1i}) = \beta_0 + \beta_1 IT_i + \beta_2 c_{1i} + \beta_3 (IT_i \times c_{1i}) + e_i \tag{3.1}$$

$$(c_{2i} - c_{1i}) = \beta_0 + \beta_1 ST_i + \beta_2 c_{1i} + \beta_3 (ST_i \times c_{1i}) + e_i \tag{3.2}$$

$$(c_{2i} - c_{1i}) = \beta_0 + \beta_1 IT_i + \beta_2 ST_i + \beta_3 (IT_i \times ST_i) + e_i \tag{3.3}$$

For each individual $i = 1, ..., n$, c_{1i} is the first-round contribution as a proxy for intrinsic motivation, and c_{2i} is the (treated) second round contribution. IT_i indicates the intervention type, ST_i the source type, and e_i the error term. We estimate model (1) for observations without source information, and model (2) separately for each intervention type. We are mainly interested in the coefficients of the interaction terms β_3, as well as in the marginal effects of either treatment IT_i or ST_i conditional on intrinsic motivation c_{1i}. Each model is estimated via OLS, using heteroskedasticity-robust standard errors.

[4] 1 subject indicated to have no education or to still be in school.

3.3 Findings

3.3.1 Descriptive statistics

Table 3.2 shows summary statistics of baseline contributions, second round contributions, their difference, as well as the difference after Falk-Kosfeld adjustment.[5]

TABLE 3.2: Descriptive statistics of contributions by experimental group

	Base contribution			Contribution			Contribution change			Contribution change > 35			n
	M	SD	Med	M	SD	Med	M	SD	Med	M	SD	Med	
Control	40.27	34.91	40	35.29	34.86	20	-4.97	23.26	0	-2.65	14.33	0	75
No Source													
Recommendation	31.59	33.68	20	31.73	29.14	35	0.14	20.37	0	-3.66	15.78	0	90
Default	30.73	30.51	20	26.66	20.89	35	-4.07	26.63	0	-8.31	20.94	0	83
Restriction	31.56	33.96	20	47.86	21.35	35	16.30	26.11	20	-0.22	18.03	0	86
Expert													
Recommendation	36.77	36.09	40	29.88	28.48	35	-6.88	19.19	0	-7.66	18.04	0	77
Default	31.77	33.42	20	27.05	22.37	35	-4.71	22.43	0	-8.22	18.72	0	73
Restriction	36.03	32.81	30	48.57	20.56	40	12.54	24.18	10	-1.11	15.28	0	79
Politician													
Recommendation	32.48	32.43	20	29.00	28.32	25	-3.48	11.10	0	-3.92	9.94	0	83
Default	33.31	30.41	30	25.81	20.66	35	-7.49	22.28	0	-8.60	17.08	0	81
Restriction	30.11	32.67	20	46.29	18.20	40	16.18	22.42	20	-0.61	11.61	0	79

Notes: Shows M = mean, SD = standard deviation, Med = Median for important variables. Contribution changes are constructed by subtracting baseline contributions from (second round) contributions. Contribution changes > 35 are constructed by setting all baseline contributions below 35 to 35 Credits, as well as all second-round contributions below 35 to 35 Credits, and then subtracting the former from the latter.

Distributions of baseline contributions do not significantly differ between experimental groups, according to a Kruskal-Wallis test ($H(9) = 6.626, p = 0.676$). This is expected because subjects were randomly allocated to treatments after making the baseline contributions.

A Wilcoxon signed rank test marginally rejects equal dependent distributions between first and second round contributions in the control group ($V = 235.5, p = 0.050$). This is evidence that subjects change their contribution in absence of experimental manipulation from round one to round two, warranting statistical analyses of difference-in-differences.

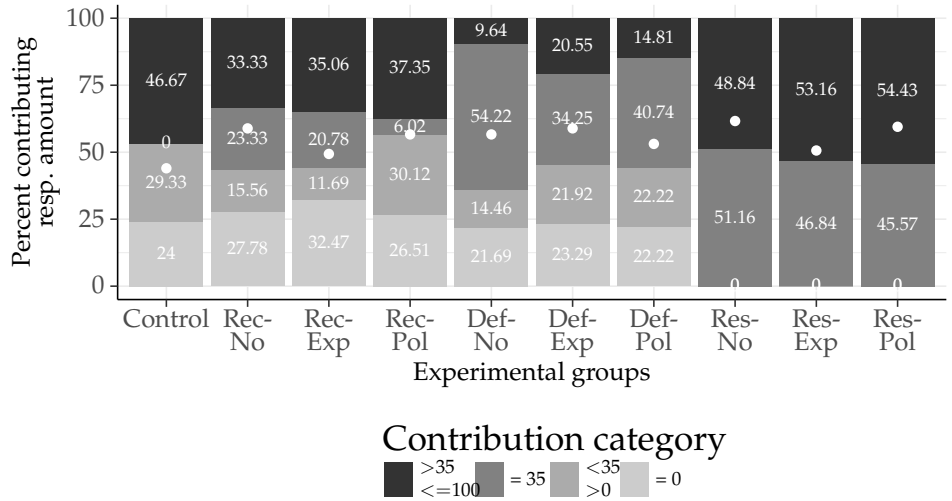
Figure 3.1 shows the distribution of contributions (c_2) after categorizing the latter into the four categories: $c_2 = 0$, $0 < c_2 < 35$, $c_2 = 35$, and $35 < c_2 \leq 100$. The white dots in Figure 3.1 show the respective fractions of subjects contributing $0 \leq c_1 < 35$, i.e. in round one.

The main insight of this graph is that the default increases the fraction of subjects contributing the default-value, while this is less pronounced in case of a recommendation. A neutral default is even more effective in attracting contributions than all three mandatory minimum contribution treatments,

[5] Because subjects in the restriction treatments had a different choice set in the second decision, i.e. a mandatory minimum contribution of 35, instead of 0 Credits, mean contributions and contribution changes are higher by design. By means of Falk-Kosfeld adjustment we can compare individual changes of contributions between both rounds that take place above 35 Credits between all treatment groups (Falk and Kosfeld, 2006).

where this value is the lowest possible value, which is consequently picked very frequently.

FIGURE 3.1: Categorized contributions by experimental group.



Notes: Shows the fraction of subjects whose contribution falls in the respective category per condition. White dots indicate the fraction of subjects that contributed $0 \leq c_1 < 35$ Credits in round one. The center of the white dot is thus the first-round equivalent to the sum of the two light shaded bar-parts depicted here. The respective fractions are shown as percentages within the respective category bars.

3.3.2 Finding 1: Effect of intervention type on contribution changes

We test for intervention effects in treatments without source information on average contribution changes, as well as on contribution changes above 35 Credits. Figure 3.2 shows the distributions and the p-values of pairwise comparisons via Mann-Whitney-U tests (p-values of all pairwise comparisons are in Appendix B.2). There is no significant recommendation-, resp. default-effect on average contribution changes. Also, the two types of interventions do not differ from each other.

In order to compare to observations facing the mandatory minimum contribution, we test for equality of distributions of contribution changes taking place *above* 35 Credits, i.e. for participants with medium to high levels of intrinsic motivation. Tests indicate that, for changes above 35 Credits, there is a default- and restriction-effect, but no recommendation-effect. Mean contribution changes above 35 Credits are negative for all treatments. Compared to the control group, defaults reduce contributions substantially (three times) more while the restriction reduces them less (one twelfth).

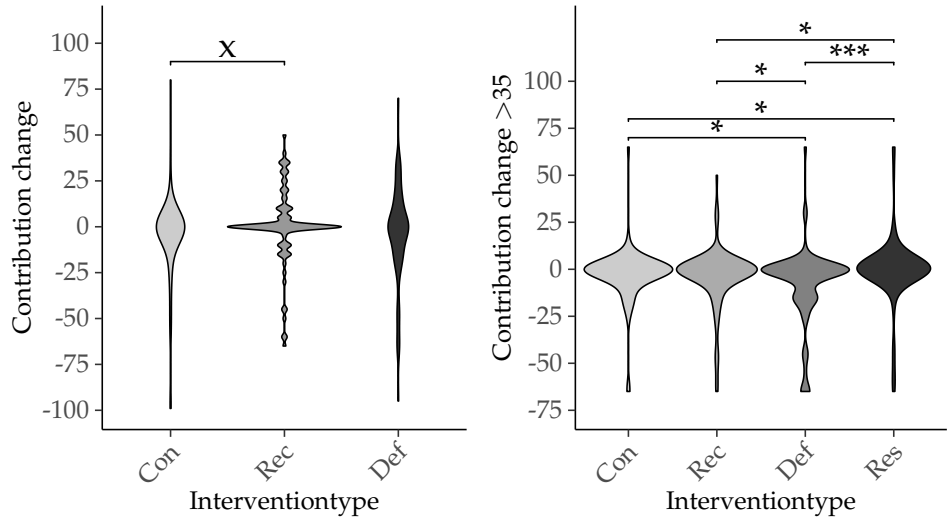
Furthermore, contribution changes above 35 also differ between intervention types. The ranking is: restriction, recommendation/control, default. The only pairwise comparison that is not significant at conventional levels is that

between recommendation and control group. Hence, relative to the control group, a default reduces and a restriction increases intrinsic motivation to contribute to the public good at hand.

F1: Neither recommendation nor default affect average contribution changes. The default and a mandatory minimum contribution decrease, respectively increase contribution changes above 35 Credits relative to the control group. The recommendation has no significant impact on contributions above 35 relative to the control group.

While there is no aggregate intervention effect for recommendation and default, indicating no support for H1a, Figure 3.1 shows that the fraction of subjects contributing exactly the recommended, resp. default value is higher compared to the control group. The absence of effects on average contribution changes is likely caused by the recommended/default value turning out to be very close to the average contribution in the first choice. In a similar experiment where the default value is 80 % of the endowment (compared to 35 % here) and the average contribution without intervention equals 18.2 % (rather than 40.3 % here), there is a significant default effect but the proportion of subjects picking the default value is much smaller (12.3 %), but also 0 % in the control group (Bruns et al., 2018). We cannot identify the level of the default as the single cause, because there are confounding differences between these experiments.

FIGURE 3.2: Kernel probability densities of contribution changes (> 35) by type of intervention



Notes: Shows kernel probability densities of contribution changes (left) and contribution changes > 35 (right) by interventions without source information. Significant pairwise comparisons from Mann-Whitney-U tests are shown. Significance levels: ^{x} ($p < 0.10$), * ($p < 0.05$), ** ($p < 0.01$), *** ($p < 0.001$).

However, the size of monetary defaults in donation contexts has been subject to investigation. For example, Goswami and Urminsky (2016) find that higher defaults lead to a small linear positive effect on donations, also reporting that lower defaults increase the likelihood of contributing a positive amount, and also that higher defaults decrease the likelihood the default value is chosen. Altmann et al. (2014) report evidence that higher defaults increase the likelihood that this value is donated, however not finding an effect on average or overall donations.[6]

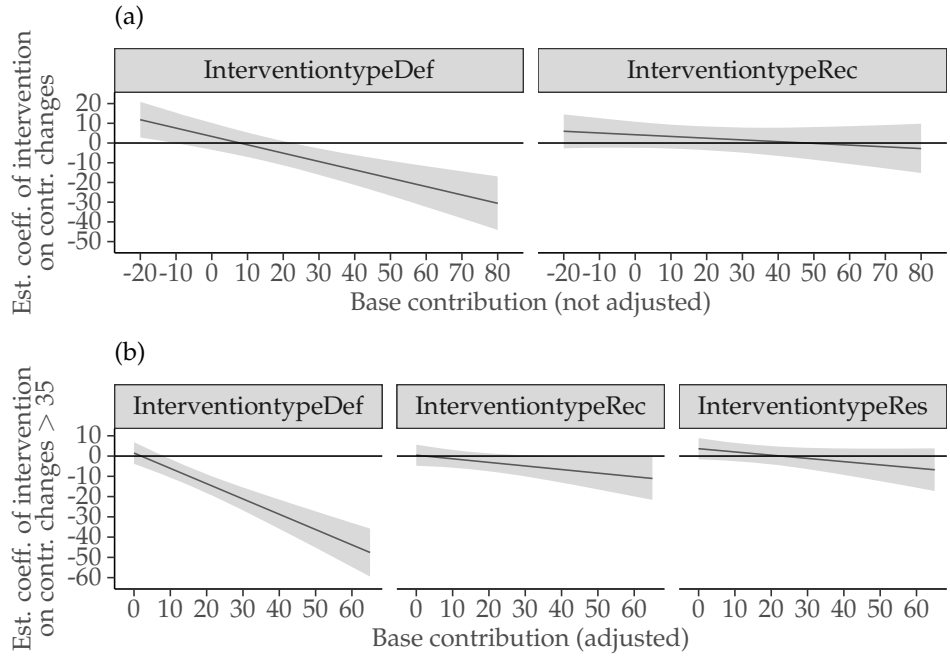
The relatively low size of the pre-defined value in the current experiment cannot explain differences between choosing rates of this value between a default and a recommendation. Subjects are more likely to contribute 35 Credits when it is set as a default, than when it is recommended. This difference helps to gain insights into likely causal channels by which the default works in this particular case. First, it is unlikely that the default works solely as a recommendation (McKenzie et al., 2006), since then we would expect a similar recommendation effect. This is in line with findings by Cappelletti et al. (2014) who find no evidence that defaults work as recommendations in a public good game. Second, the same might be the case for the default working due to providing a reference value or anchor, associated with loss aversion (Samuelson and Zeckhauser, 1988; Dinner et al., 2011; Dhingra et al., 2012). Both interventions can be said to provide a reference value or anchor, although arguably a recommendation does this not as prominently as a default. However, both mechanisms might play a larger role for subjects who are insecure about their intrinsic motivation and are consequently influenced by this value (Jachimowicz et al., 2017). Third, it is likely that the defaults works due to inertia (Johnson and Goldstein, 2003), i.e. peoples' tendency to stick to it to reduce effort. Inertia is consistent with the observed neutral aggregate effect, assuming that it is independent of intrinsic motivation. Subjects with low, as well as high intrinsic motivation stick to the default, and thus increase/decrease their contribution, canceling out on the aggregate. We further investigate this in Section 4.4. While this is potential evidence for an inertia explanation of the default, the fact that this is also the case for a recommendation suggests that anchoring is also possible. Note that the attraction effect of the default cannot be caused by participants economizing on efforts in making a first-time choice in an unfamiliar context (unlike in Bruns et al. (2018)), since they have already made such a choice a few seconds earlier.

3.3.3 Finding 2: Interaction of intervention type and intrinsic motivation

The experimental design allows to test for treatment effects conditional on intrinsic motivation, providing us with evidence regarding an intervention's relative potential to crowd in or crowd out intrinsic motivation. Panel (a) of Figure 3.3 plots the estimated effects on contribution changes caused by the

[6] In both of the previously mentioned studies, the respective design does not allow to translate the default size to a fraction of the endowment, because the latter is unobserved.

FIGURE 3.3: Intervention effects on contribution changes (> 35) relative to control conditional on baseline contributions



Notes: Shows estimated effects of using an intervention relative to no intervention conditional on base contributions on contribution changes (top), as well as on contribution changes > 35 conditional on baseline contributions > 35. (Adjusted) baseline contributions are median-centered. Grey areas represent simulated 95 % confidence intervals (10,000 draws).

neutral default, resp. the recommendation relative to no intervention, conditional on median-centered base contributions from model (1). Estimation results are shown in Table B.5.[7] Defaults increase contributions relative to base contributions for subjects with intrinsic motivation below the median, and decreases them for subjects with higher-than-median intrinsic motivation. Note that the target level of all interventions is close to the median contribution in choice 1 that is used to quantify pre-existing intrinsic motivation. Both effects cancel out on the aggregate, consistent with the non-significant difference between distributions in Def-No and the control group (Finding 1). The estimated coefficients are not significantly different from zero for recommendations. The negative marginal effect of intrinsic motivation on contribution changes in the control group gets stronger when the default (instead of no intervention) is used to influence contributions. The marginal effect in the setting with the anonymous recommendation is not significantly different from the control situation. Post-hoc tests indicate that the marginal effect of intrinsic motivation is significantly lower in case of a default compared to a recommendation ($B = -0.34, CI_{95}[-0.54, -0.14], p = 0.001$).

[7] Relevant estimates are robust to the inclusion of covariates and can be obtained from the authors upon request.

Panel (b) of Figure 3.3 shows estimated coefficients of adjusted contribution changes caused by the neutral default, recommendation, and restriction, relative to no intervention, conditional on centered and adjusted base contributions. Graphs show that when subjects face a default they reduce their contribution (typically by choosing the default). The effect is of course more drastic for high base contributions. The effect is very small for recommendations and non-existent for restrictions. The marginal effect of intrinsic motivation on adjusted contributions in the control does not change due to the restriction or the recommendation, albeit due to the default. Post hoc tests indicate that the difference of the marginal effect is significantly lower for the default compared to the recommendation ($B = -0.58, CI_{95}[-0.84, -0.32], p < 0.001$), and compared to the mandatory minimum contribution ($B = -0.60, CI_{95}[0.32, 0.87], p < 0.001$). The effect of intrinsic motivation does not differ between the mandatory minimum contribution and the recommendation ($B = 0.014, CI_{95}[-0.32, 0.35], p = 0.933$).

F2: Both the mandatory minimum contribution and the recommendation do not crowd out high levels of intrinsic motivation to contribute to climate protection. The default crowds out high intrinsic motivation.

Note that, although the default also crowds out, this effect is offset in the aggregate by a positive (crowding-in) effect for subjects with low intrinsic motivation. Put differently, contributors that would have contributed more than the default value are reducing their contribution, while contributors that would have contributed less than the default value are increasing their contribution. This effect appears to be exclusive to the default, as there is no such significant indication for a recommendation or mandatory minimum contribution. The fact that we observe motivation crowding out of highly intrinsically motivated individuals by a default is consistent with findings reported by Goswami and Urminsky (2016). Interestingly, in our setting, enforcement of a minimum contribution did not reduce individual contributions made by those who had previously been contributing above the regulatory requirement. Thus, our findings do align with Falk and Kosfeld (2006) and Goeschl and Perino (2012) who both find no crowding out for restrictions originating from a neutral experimenter. Along the line of reasoning of Reeson and Tisdell (2008) this may have been caused by subjects with high intrinsic motivation perceiving the mandatory minimum contribution as a means to coerce free-riders to conform to a perceived social expectation of contributing to climate protection. In this case, a retraction of this intervention may lead to negative consequences like motivation crowding only in the long-run, because this would then be perceived as a retraction to the previous "unfair" system.

3.3.4 Finding 3: Effect of source type on contribution changes

To see whether the type of source affects contributions to climate protection, Figure 3.4 shows the distributions of contribution changes for the different interventions, disaggregated by the source type. Compared to the recommendation without source information, providing political, resp. expert source information induces a significantly stronger reduction in contributions. However, all recommendation treatments are indistinguishable from the control group. Relative to a neutral default, source information has no significant effect. Again, distributions in all default treatments are indistinguishable from the control group.

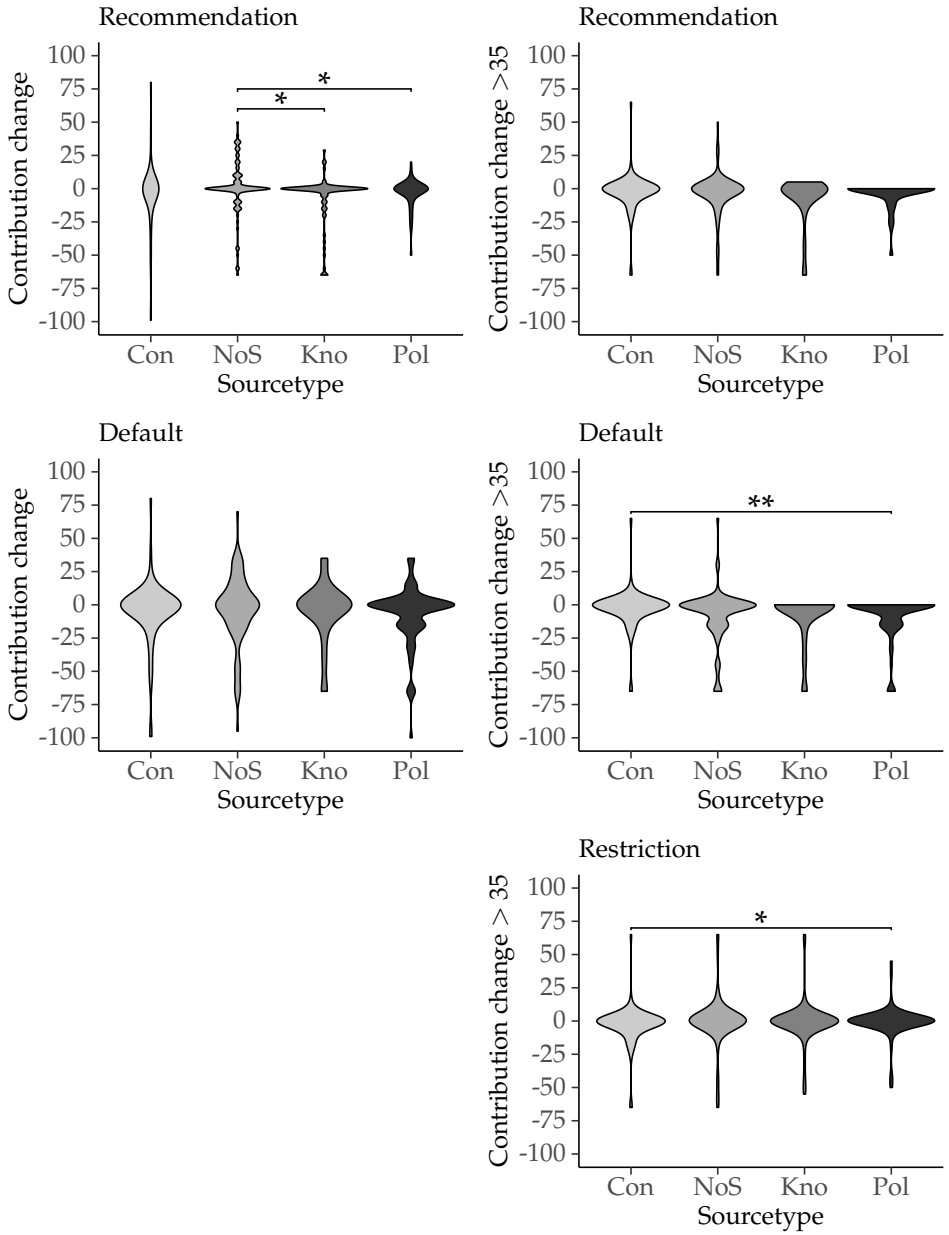
The null hypotheses that the distributions of contribution changes above 35 Credits are equal across source types is not rejected for either type of intervention.

Neither the recommendation with any type of source information leads to adjusted contribution change distributions different from the control, nor do combinations of expert information accompanying the default. Compared to the control group a default and a restriction imposed by a politician significantly affect contribution changes above 35. The default induces a stronger reduction and the mandatory minimum a more moderate reduction in contributions relative to the case without an intervention. P-values of all pairwise comparisons are in Appendix B.2.

F3: Neither the recommendation nor the default with source information affect contribution changes compared to the control group without an intervention. While providing source information has no effect for the default, it decreases the contributions for the recommendation relative to an identical intervention without a specified source. Changes in contributions above 35 are affected if a politician is identified as the source of the intervention. The decrease in contributions is larger for the default and smaller for the restriction, each compared to the control group.

There is no evidence for an effect of source information provision for a default on average contribution changes. However, a default decreases high contributions when initiated by a politician. This can consequently be explained by the politician being perceived as profit-maximizing (Brown and Krishna, 2004) or an information asymmetry about the social preferences of the decision makers relative to the politician (Altmann et al., 2015). According to Kuang et al. (2007) the finding that a recommendation decreases contributions only when complemented by source information suggests that both an expert and a politician might be perceived as self-interested by the decision maker. However, this does somewhat contradict the positive effect of political source information on the mandatory minimum contribution. A post-hoc rationalization of this finding could be that politicians have a mandate to coerce people into contributing to climate protection, but that they should not use a default, which could be viewed as more covert relative to a mandate, to get people to contribute.

FIGURE 3.4: Kernel probability densities of contribution changes (> 35) by type of source



Notes: Shows kernel probability densities ofcontribution changes (left) and contribution changes > 35 (right) by source information for each intervention. Significant pairwise comparisons from Mann-Whitney-U tests are shown. Significance levels: ^{x} ($p < 0.10$), * ($p < 0.05$), ** ($p < 0.01$), *** ($p < 0.001$).

3.3.5 Finding 4: Interaction of source type and intrinsic motivation

We also test for effects of the source type conditional on intrinsic motivation, thus investigating the motivation crowding potential of this aspect of the decision context. Panel (a) of Figure 3.5 plots estimated coefficients of contribution changes caused by providing source type information along different levels of baseline contributions, from model (2). Estimation results from model (2) are shown in Table B.6.[8] Relative to the negative effect of baseline contributions for recommendations without source information, provision of expert and political source information has no significant effect. The positive baseline contribution effect for a default also does not change significantly due to provision of source information.

Panel (b) shows that the effects of complementing a recommendation with information about an expert or politician do not significantly change the effect of adjusted baseline contributions on contributions above 35. The conditional effects of providing source information for a default on contributions changes above 35 are primarily positive for most adjusted baseline contributions. However, the respective interaction coefficients are not significantly changing the adjusted baseline contribution effect. The positive conditional effects for high adjusted baseline contributions suggest that the crowding out effect of the default is dampened when an expert or a politician are identified as the source of the default. For restrictions, source information does not have an interaction effect with baseline contributions.

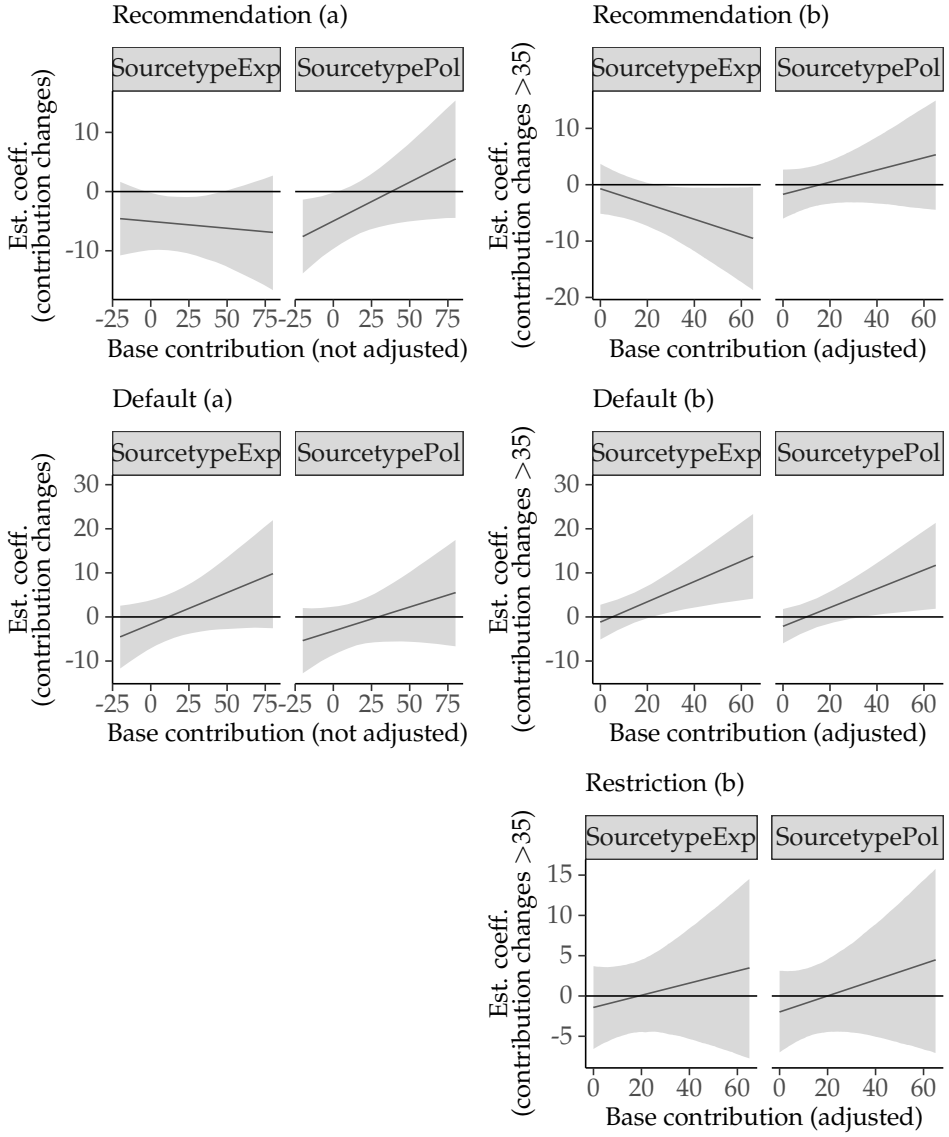
The effect of intrinsic motivation given an intervention does not depend on whether the source is an expert or politician, as opposed to an unspecified source. This is the case for adjusted as well as unadjusted contribution changes as the dependent variable.

F4: The provision of information about the source does not crowd out decision makers' intrinsic motivation to contribute to climate protection.

Although not confirming H2b, the absence of any significant source effect on motivation crowding complements results in the literature. Falk and Kosfeld (2006) find that control aversion only occurs when subjects see the intervention as a conscious decision of an actor that is trying to control them in order to serve their own self-interest but not so if the intervention originates from the experimenter. Our results are in line with theirs if both sources are not perceived as (narrowly) self-interested. This is plausible, since the person identified as a source does not receive any monetary or other direct benefits as a result of the choices made in this experiment - unlike the sources in their experiment. An important insight here is that the neutral default appears to be the major cause of negative motivation crowding, relative to a default implemented by an expert, resp. a politician.

[8] Relevant estimates are robust to the inclusion of covariates and can be obtained from the authors upon request.

FIGURE 3.5: Source effects on contribution changes (> 35) relative to no source information conditional on baseline contributions (> 35) for different intervention types



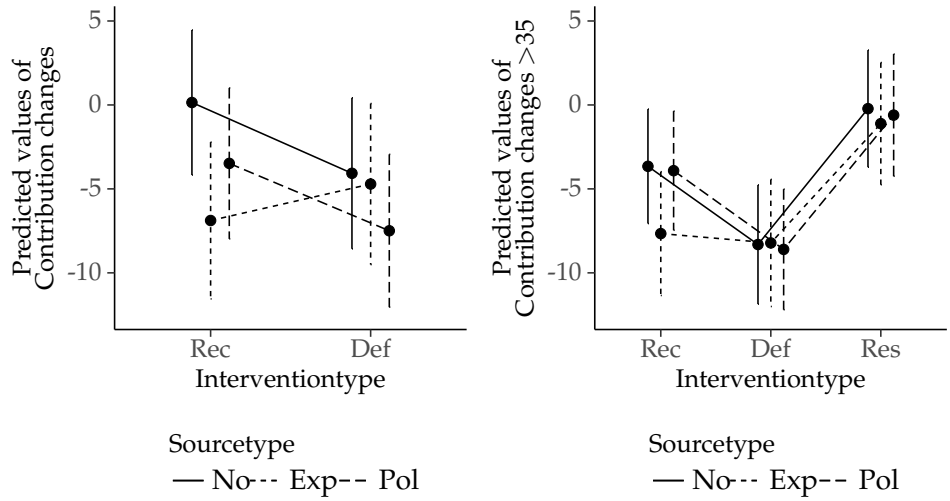
Notes: Shows estimated effects of providing source information relative to no source information for the respective intervention and conditional on base contributions on contribution changes (left), as well as on contribution changes > 35 conditional on baseline contributions 35 (right). Baseline contributions (> 35) are median-centered. Grey areas represent simulated 95 % confidence intervals (10,000 draws).

3.3.6 Finding 5: Interaction of intervention- and source type

Estimation results from model (3), i.e. estimates of OLS-regression including an interaction of intervention and source type are presented in Table B.7.[9] None of the estimates of the respective interactions are statistically significantly different from zero at conventional levels (all $p > 0.2$). This suggests that neither the impact of a recommendation nor restriction relative to a default, on individual contribution changes differs conditional on the type of source responsible for the implementation of the respective intervention. Additionally, Figure 3.6 shows that the differences of the intervention effects on predicted (adjusted) contribution changes do not change significantly when varying the source. Overall, there is no support for H2c.

F5: The influence of an intervention on contributions to climate protection does not depend on the type of regulator responsible for its implementation.

FIGURE 3.6: Predicted values of contribution changes (> 35) for source type conditional on intervention type



Notes: Shows values for contributions changes (left) and contribution changes > 35 (right) predicted by the OLS-models in Table B.7 conditional on source type and intervention type.

In line with findings 2 and 4, where source effects were minor or non-existent, there is no evidence of a substantial interaction effect between interventions and sources responsible for the intervention. In relation to results in the literature, this could imply that the source type variation used in this experiment was missing an aspect that was crucial for producing significant differences in responses. One candidate is that the payoff of the source was not affected by choices made within this experiment, which is what seems to be driving the source effects in Falk and Kosfeld (2006). The source specific differences reported in Goeschl and Perino (2012) and Perino et al. (2014)

[9] Relevant estimates are robust to the inclusion of covariates and can be obtained from the authors upon request.

occur with monetary incentives, not with nudges or restrictions. Moreover, on top of the source, these earlier experiment simultaneously varied whether the change in conditions was framed as an explicit intervention as opposed to a change e.g. in market conditions.

3.4 Discussion

We critically reflect on and discuss two aspects of the experimental design that might inhibit the potential of this experiment to make unbiased causal inferences: First, a within-subject design bears the potential of carryover effects, i.e. interpreting the measurement of intrinsic motivation as a treatment hypothetically affects second round contributions, and also may do so differently by treatment. To address these issues, we conducted one-round versions of two experimental groups (Control and Def-Exp) and tested if measuring baseline contributions prior to random treatment allocation significantly affected second round contributions compared to a between-subjects design without baseline measurement. We find that the distributions of second-round contributions in both two-round-designs were not significantly different from respective distributions in the one-round-designs (Control-1R vs. Control-2R: $W = 2896.5, p = 0.174$; Def-Exp-1R vs. Def-Exp-2R: $W = 2791.5, p = 0.717$). Additionally, a robust OLS-regression of contributions on the number of rounds interacted with the treatment does not provide a significant interaction effect ($B = -7.44, p = 0.239$) (see Table B.8 in Appendix B.2).

We also test whether announcing to subjects ex-ante that they were about to make two decisions systematically affected their base contributions. Testing for equal contribution-distributions in the one-round control group (no announcement of a second contribution decision) versus pooled baseline contributions from control and Def-Exp groups (including announcement of a second contribution decision) indicates marginal significance ($W = 5648, p = 0.071$). This could suggest that subjects initially contribute more when they are allowed to change their decision at a later stage, compared to when they only make one decision. This may serve as evidence that subjects "overstate" their intrinsic motivation in our experiment. The difference-in-difference approach applied in our analysis should eliminate any distortion caused by this phenomenon.

Second, sample selection caused by attrition (attrition bias) potentially causes biased results (see e.g. Gerber and Green, 2012). Although (potential) participants are reminded to participate several times, 143 (chose to) drop out at either point during the trial. Tests reported in Appendix B.2 indicate that subjects who drop out of the experiment are older on average. Data do not suggest that dropping out of the experiment correlates with treatment assignment ($\chi^2(11) = 13.428, p = 0.266$). The latter fits the observation that most subjects dropped out before they had to make the first contribution decision, i.e. before they encountered the treatment. This suggests that subjects do *not* drop out because of the treatment. Nevertheless, we still do not know what subjects that dropped out would have contributed.

3.5 Conclusion

This paper presents framed field experimental evidence on how different types of interventions, as well as information provided on regulators responsible for such interventions affect the relation between intrinsic motivation and individual voluntary contributions to climate protection. Findings show that *neither* the recommendation, *nor* the default affect *average* contributions to climate protection if interventions target a contribution level close to the pre-intervention average. However, while the default decreases high contribution levels, the mandatory minimum contribution does the opposite, and the recommendation has no such effect, relative to a control group that was not exposed to an intervention. While at the same time neither the recommendation, nor the default with source information affect average contribution changes, the recommendation without information is more effective than the recommendation from an expert or politician. The mandatory minimum contribution increases high contributions when initiated by a politician but not when initiated by an expert, whereas the default decreases high contributions when initiated by a politician but not when initiated by an expert. Contrary to our expectation, data do not indicate that the mandatory minimum contribution crowds out contributions. Affirming our expectation, neither does the recommendation. However, there is evidence that the default crowds out contributions for subjects with high intrinsic motivation. The provision of regulator information, however, does not interact with intrinsic motivation. The relative effects of interventions do not change depending on the regulator information provided.

Taken together, these findings suggest that defaults, when set relatively low, can crowd out highly motivated contributors, contrary to recommendations and mandatory minimum contribution levels. Additionally, this process is a potential explanation for the absence of a default effect that has been found in earlier contributions. This experiment clearly shows that the default is effective in raising the likelihood that the target level is chosen. Since this level by coincidence turned out to be close to the pre-intervention level of contributions, no change in average contributions was observed. The absence of a recommendation effect, however, is likely to be due to the generally low impact of the intervention. Across the three versions of the recommendation treatment used in this experiment, the changes in contributions was minimal both compared to the no-intervention group and pre-intervention levels.

Future research could test whether relative effectiveness of the interventions discussed here is conditional on the promoted values in relation to intrinsic motivation. A higher relative value, for example 80 Credits, might lead to different conclusions with regards to effectiveness and interaction effects. Furthermore, since we hypothesize that the crowding-out effect of the default might be due to it working as a tool for effort-reduction, it would be interesting to investigate whether one would observe similar effects when the default, or any other type of nudge, does strictly not impose effort on the decision maker who wants to deviate from the value. However, this might be

difficult to accomplish since cognitive costs might always be imposed when subjects need to consider promoted values that deviate from their intrinsic motivation, e.g. because they view the intervention as a hint to a socially desirable contribution.

Findings presented here compare a recent and prominent policy-type, a nudge, to more established policies, i.e. recommendations as pointers, and mandatory minimum contributions as pushes, within the context of private public good provision. This comparison is especially relevant to the ongoing discussion about aspects of autonomy of so-called "libertarian paternalistic" policies (Sunstein, 2016; Schubert, 2017) used to internalize negative external effects (i.e. to reduce free-riding), as well as to the established discussion of motivation crowding of economic incentives. Furthermore, it adds to the literature comparing the cost-effectiveness of behavioral to conventional interventions (Benartzi et al., 2017).

4 Does proclaimed doubt in media spill over to doubt in science? A laboratory experiment in the context of climate change

Author: Hendrik Bruns
The data and scripts in support of the findings are openly available at OSF under DOI 10.17605/OSF.IO/86SJX.

Abstract: Labeling news as fake is a recent phenomenon occurring predominantly online, and increasingly in political online environments. This paper investigates the influence of proclaimed doubt in media independence on trust in news- and scientific reports on climate change. Evidence from a preregistered laboratory experiment does not suggest that reading a media-critic statement affects perceived trust in the media-, or the scientific source. Bayesian analyses provide a practical interpretation of the null findings, and further analyses show that the proclamation decreases trust in the scientific source when subjects read the media article first. Findings add to the emerging literature on fake news, echo chambers and filter bubbles, suggesting that labeling stories or outlets as fake news may not affect public opinion. Further research is needed to substantiate this conclusion.

4.1 Introduction

Trust and confidence in the mass media among US-Americans has been dropping steadily from 53% in 1997, to 32% in 2016. (Swift, 2016). At the same time, scientists observe distrust in the natural and social sciences dealing with climate change, particularly in the causal role of humans (Tollefson, 2010; Dunlap, 2013; Boussalis and Coan, 2016). This development is striking because there is broad scientific agreement that humans contribute to climate change (Weber and Stern, 2011; Cook et al., 2013; Cook et al., 2016). Extant literature investigates potential causes for increasing distrust in the science of climate change (Hmielowski et al., 2014) or environmental degradation (Zhou, 2014), as well as potential repercussions of such developments (Feldman et al., 2014). However, until now there is no systematic analysis of the causal influence of statements by "important persons" publicly proclaiming doubt in the independence of news media on public perceptions of the news media, as well as on public perceptions of scientific topics regularly discussed in the news, such as climate change.

What is meant in this context by an "important person" can be exemplified by US-President Donald Trump, who publicly engages in official "media-bashing"[1]. This behavior is characterized partly by his increasing usage of the term fake news as a label for certain news media outlets or news stories, on his official Twitter account (see Appendix C.2 for Panel A in Figure C.2). The fact that he does so in up to 9.5% of his tweets (December 2017) which get read and discussed extensively (Robinson, 2017), potentially increases public distrust in the news media. Of course, public discussion of his statements concerning fake news include criticism as well as endorsement, but they appear to have the potential to motivate (a sub group of) people to strengthen their beliefs and attitudes if they correspond to those expressed by Donald Trump. Similarly, statements and actions by the President of Turkey, Recep Tayyip Erdoğan on and against allegedly dependent media outlets (Waldman and Caliskan, 2017) have fueled discussion of the impact of such statements on the general populations' attitudes and skepticism towards news media (Daily Sabah, 2017).

Public denunciation of news media is often motivated by partisan attitudes. For example, Donald Trump primarily labels democratic-oriented news outlets as fake news, because they predominantly tend to oppose his political views. These outlets often take a particularly acknowledging stance on the occurrence of manmade climate change, a topic towards which Donald Trump is skeptical. This is indicated by his alleged ban of the term "climate change" in the US Department of Agriculture (Milman, 2017), the election of an alleged climate change denier, Scott Pruitt, as the head of the Environmental Protection Agency (Davenport and Lipton, 2016), as well as the United States withdrawal from the Paris Agreement (Shear, 2017). Consequently, labeling a subset of news outlets as fake news may decrease the

[1] Here defined as a harsh, usually derogative and dismissive form of referring to news articles, or news outlets.

trust in these outlets, and potentially the trust in the science underlying climate change, which they report. The fact that he uses predominantly negative wording, according to the AFINN sentiment score (Nielsen, 2011), in tweets referring to fake news highlights the importance and emotionality of this topic (see Appendix C.2 for Panel B in Figure C.2).

This paper asks whether increasing the salience of doubt in *media*-independence causes a short-term reduction of the perceived source trustworthiness of a *news media* and a *scientific* article on climate change. Evidence comes from a laboratory experiment, in which subjects read two largely identical articles from a news- resp. scientific source, on recent events in climate change. After having read each text, respondents indicate their credibility-ratings of the respective source and content of the article. Participants in the treatment group do so after having read a non-fictional sentence challenging the independence of news media.

Findings do not suggest that perceptions of the media, or the scientific source change due to a primed elicitation of distrust in media by means of exposition to a sentence expressing doubt in media independence. The same holds for other outcome variables measuring perceived source trust, source expertise, and content credibility. Further analyses of heterogeneous treatment effects reveal a significant interaction between the treatment and order in which participants were exposed to the news articles.

The study contributes to the recent literature on fake news, echo chambers and filter bubbles by investigating an additional potential explanation for the contemporary increase in media- and science-skepticism, especially in the context of man-made climate change. It also provides an important vantage point for further analyses that may focus on the role of social networks and partisan source-characteristics on doubt proclamations.

The article is structured as follows: Section 4.2 will outline the relevant literature and hypotheses. Section 4.3 describes the laboratory experiment. Section 4.4 provides descriptive and inferential findings, while Section 4.5 discusses and concludes.

4.2 Relevant literature and hypotheses

This section outlines literature relevant to answer two questions: First, why could reading a statement expressing doubt in the independence of news media affect one's trust in a specific news media source? Second, why could such an expression of doubt in media- independence affect one's trust in a specific scientific source?

The two-step flow model (c.p. Katz et al., 2005) explains the procurement of information as a process of information traveling through media, from "opinion leaders" using such media (step 1) to less engaged "opinion followers" (step 2). Although this model was built way before the dawn of online social networks, or even the internet, its predictions are consistent with contemporary observations in these environments. Recent research indicates that a person, especially one with high influence via social network paths, i.e. high centrality, can exert profound influence on how news is processed

and their credibility is perceived by other people in their network. More precisely, Turcotte et al. (2015) find that, the more a friend is perceived as an opinion leader, the higher the trust in the news media outlet responsible for the article shared by the friend. Also, consumers indicated to be more willing to seek future information of a person they viewed as an opinion leader. Perceptions of likewise received news can interact with the receiver's attitude towards and perceptions of the sender, using these as heuristic cues in order to assess the information provided by the sender (Chaiken, 1980). Together, these processes predict that the majority of opinion followers get trusted information from leaders of whom they have a positive attitude because of the proximity in their social network. This process can create "echo chambers", referring to a framework of positive feedback mechanisms that mutually reenforce existing opinions, especially in the context of climate change (Walter et al., 2018).

Additionally, cognitive heuristics simplifying the processing of information, such as confirmation bias, ingroup bias, and negativity bias, are hypothesized to lead to an increasing tendency of consumers to reduce their personal set of trusted news outlets to those echoing their pre-existing beliefs and attitudes (Knobloch-Westerwick et al., 2015; Knobloch-Westerwick et al., 2017).[2] The existence of confirmation bias, defined as selective exposure to and interpretation of information in alignment with pre-existing attitudes (Nickerson, 1998), is especially interesting in this respect, given that it may lead to a biased information environment created by the consumer, in response to a perceived bias in or independence of news media coverage. Extreme and solidified attitudes and misinterpretation of facts about current events, as well as the false consensus effect, which suggests that people overestimate the degree to which their believes, opinions, etc. are representative of those of others, are potential negative outcomes of such a process (Stroud, 2008; Leviston et al., 2013). Repercussions of confirmation bias are intensified by contemporary possibilities to choose attitude-consistent news from an abundant and diverse set of online media and social networks, catering to all beliefs and attitudes. Mechanisms on social network sites, such as Facebook, contribute to this with algorithms determining the order in which users are exposed to articles in their News Feed, based, for example, on past behavior, and a user's social network (Bakshy et al., 2015; Knobloch-Westerwick et al., 2017; Allcott and Gentzkow, 2017).

Such a selective exposition to and consumption of news, in combination with an expressive dissemination of distrust in the independence of news media by focal people, defined as central nodes having many (outgoing) paths in a social network, could interact, thus amplifying the false consensus effect. This interaction potentially causes a non-negligible reduction of consumers' trust in certain news media outlets, especially those propagating views contradicting their own.

[2] Ingroup bias describes information procurement as an outcome of social identity attributions, whereas negativity bias predicts that negative news get more attention than positive news (Knobloch-Westerwick et al., 2017).

While less trust in certain news media outlets, as well as reservation regarding their independence may be perfectly adequate, both states bear potential for negative consequences. These consequences are worthy of scientific investigation since public attitudes towards the credibility of news, and the media in general can influence public opinion on policy matters (Page et al., 1987). Consequently, discrediting the independence or credibility of news media may constitute a promising strategic lever for politicians or other policy-makers to generate, or steer public support of their policies.

How does this relate to trust in science? For many people, (online) news media are an important source of information on scientific discoveries, especially related to the science on climate change (Schäfer and Schlichting, 2014). For example, the website "Climate Debate Daily"[3] links to more than 5,000 articles published between 2008 and 2016, supporting or doubting the existence of global climate change. There are many more to be found on the internet. While such an abundance of media articles is important for the formation of opinions, it also bears potential risks. News media often make scientific discoveries, e.g. those on climate change, easier to comprehend for non-experts (Moser, 2010). However, since the distrust in the news media increases, and their independence may further be questioned publicly by influential "opinion leaders", not just credibility in the news media, but also in the topics they are writing about, and ultimately also the sources of those topics, i.e. climate science itself, may lose credibility as perceived by parts of the public. A link between "opinion leaders" that discredit news media, and ultimately spur doubt in the underlying science could explain why public trust in climate science stagnates or decreases.

A conceptual model outlining this process analytically predicts that, stated disbelief in the independence of news media reporting scientific results on climate change might induce consumers to generalize such distrust to the content (climate change), as well as to the source of the content (science), for reasons of consistency. In other words, the content and source of a scientific climate change article might suffer in its perceived credibility and trustworthiness because of externally motivated distrust in the independence of the news media source reporting the same, or similar evidence.[4]

This article proposes that exposition to expressed distrust in the independence of news media primes suspicion of the news source. Furthermore, it proposes that, even though such a statement explicitly voices suspicion of the media because of its alleged lack of independence, a motivation for consistency leads consumers to experience an increased general skepticism and consequently lower trust in a comparable article that, however, has been

[3] Accessible here: `www.climatedebatedaily.com`.

[4] The solution aversion model can serve for an allegory: solution aversion, which is a type of motivated skepticism, explains the tendency of people to negate the existence of a problem because they do not support the policies that are primarily advocated to solve the problem (Campbell and Kay, 2014). Similarly, people might negate a scientific source reporting a problem because they do not support the source of a news article reporting the same problem.

published by a scientific source. From the above reasoning follow two hypotheses:

H1: Increasing the salience of doubt in the independence of media reduces perceived trustworthiness of the source of the media article on climate change.

H2: Increasing the salience of doubt in the independence of media reduces perceived trustworthiness of the source of the scientific article on climate change.

Note that the experiment described in the remainder of this article is not designed to differentiate between the potential causal underpinnings of the empirical regularities proposed in the two hypotheses. This is further elaborated on in the discussion section of the paper.

4.3 Experiment

The experimental paradigm randomly allocates subjects to treatment before they read two articles. After having read each of them, information on dependent variables is gathered by asking subjects to indicate their level of agreement with adjectives describing the content and source of the articles. Central parts of the experimental design and statistical analyses were pre-registered via AsPredicted on 22 August 2017 (#5219) and made publicly available on 27 February 2018. The pre-registration document is in Appendix C.3.[5] As outlined in the pre-analysis plan, an a priori power analysis was conducted in order to decide on the number of participants. Since I am not aware of similar experiments that would allow to specify an expected standardized effect size, I decided on a minimum power of 80 % in order to detect a medium standardized effect size Cohen's $d = 0.4$, given a conventional alpha level of 5 %, an independent samples Wilcoxon Mann-Whitney U test, and a *one-sided* hypothesis.[6]

4.3.1 Experimental design and manipulation

Subjects, irrespective of treatment assignment, read two articles, the order of which varied randomly between subjects. One article was taken from the online portal of the German newspaper outlet ZEIT ONLINE (Zeit Online, 2016), while the other was taken from the website of the National Oceanic Atmospheric Administration (NOAA) (National Oceanic and Atmospheric Administration, 2016). The newspaper article is largely based on the NOAA article. The topic of both articles is climate change, which is a relevant topic

[5] The report can also be accessed online: `https://aspredicted.org/w98wj.pdf`.

[6] Researchers often abstain from or criticize the use of one-sided tests because it might indicate that they were used to turn non-significant results significant by dividing the p-value by two and accordingly re-formulate the hypotheses, see for example Cho and Abe (2013). This inflates the alpha error. Since this analysis is pre-registered, this is not a problem here.

with respect to the interplay between media, science, and external influence on credibility perceptions (Dunlap, 2013). Both texts, translated and in the respective original language, are in Appendix C.1. The NOAA article was translated into German for the experiment. After reading each article, subjects indicate their credibility ratings of the source and content for the respective article. Subjects are randomly allocated to the treatment. Participants in the treatment group read a quote taken from an interview of the newspaper DIE ZEIT with the Turkish president Recep Tayyip Erdoğan:

"I don't believe there is such a thing as 'independent media' anywhere in the world. At some level, they are all – whether print or broadcast media – dependent, either ideologically, or they are pursuing their own interests. If there were such a thing as independent media, we wouldn't have all these problems. We see things quite clearly: They head in whichever direction the wind is blowing. The German media is no different. Nobody can say that isn't the case. We know very well that's how things are." (Di Lorenzo, 2017)

At this point, participants were not told that the quote was from the Turkish president. They were told after the experiment. This design aspect is discussed in the remainder of this article.

4.3.2 Outcome measures

The primary outcomes measure subjects' perceived trustworthiness of the source of either article. These variables consist of responses to a 5-point semantic differential item that is part of a scale used to measure celebrity endorsers' perceived expertise, trustworthiness, and attractiveness (Ohanian, 1990). Subjects also indicated their perceptions regarding the sub-scales on perceived source expertise and trustworthiness, which are secondary outcome measures. A further secondary outcome is perceived content credibility. This construct is measured via four 5-point semantic differential items of a content credibility scale based on Gaziano and McGrath (1986). Additionally, participants in the treatment group indicate their level of agreement with the treatment-quote on a 5-point Likert scale. The questionnaire is in Appendix C.3.

4.3.3 Participants

Participants are drawn from a pool primarily consisting of undergraduate students from the University of Hamburg, Germany. All potential subjects voluntarily signed up to participate in laboratory experiments. They are invited to experiments conducted in the lab regularly and are free to choose whether they want to participate, or not. The 194 participants[7] that participated in this experiment have a mean age of 24.81 years (SD = 5.02, Median = 24), 60.82 % are female, and 28.35 % are from the faculty of economic and

[7] 2 observations were dropped from further analysis, because the subjects took less than two standard deviations than the mean duration to read one of the news articles.

social sciences, 14.95 % social sciences, 13.4 % from management, 6.19 % history, 5.67 % law, 1.03 % philosophy, 30.41 % others.

4.3.4 Procedure

Experimental sessions took place between 23 August 2017 and 6 November 2017. Subjects participated in the experiment after participating in another, quite short, experiment in which they were endowed with money and then asked to decide if they wanted to donate any amount to climate protection (see Bruns et al., 2018). After filling out a post-experimental questionnaire from the previous experiment, subjects were kindly asked to participate in the follow-up experiment, while their private payments were prepared. Participants did not get any financial compensation for participating in the second experiment. However, their expected payouts in the first experiments were higher than usual for this subject pool. While they could earn up to 10 Euro ($ 11.89) in the experiment, which lasted up to 30 minutes, subjects from this pool normally expect to earn on average 10 Euro for 60 minutes. The second experiment was not announced to subjects before the first experiment ended. Subjects read instructions on the computer screen in front of them, before being able to proceed in their own pace. After reading each of the two articles, and possibly the treatment manipulation, subjects had to conform that they read and understood the text before they could proceed to the next stage. This was implemented in order to increase the likelihood that subjects fully comprehended what they were reading, and to remind them that thorough reading and comprehension was important. Subjects in the treatment group, after having read the quote, indicated their level of approval of it. After having read either article in random order, subjects provided their respective content and source credibility ratings. At the end, subjects were briefed about the origin of the treatment sentence they read at the beginning. They were informed that it was a quote from the Turkish President Recep Tayyip Erdoğan from an interview with the German newspaper outlet DIE ZEIT from the 5th of July, 2017. Participants were located in small booths in a laboratory. Visual, as well as verbal communication between participants was not allowed and/or possible during the experiment. The study was conducted using the experimental software z-Tree (Fischbacher, 2007). Participants were recruited via hroot (Bock et al., 2014).

4.3.5 Statistical analyses

The statistical procedures outlined in the pre-analysis plan were conducted to test the hypotheses. The treatment effects on the two main outcome variables *trustworthiness of media source* (H1) and *trustworthiness of scientific source* (H2) are tested with one-sided Chi-squared tests, Mann-Whitney tests, as well as ordinal logistic regression. Additionally, Bayesian ordinal logistic regression (Gelman et al., 2014; Kruschke, 2015) is used in order to quantify the relative evidence of the hypothesis that there is a treatment effect vs. no effect, resp. whether the treatment effect is negative vs. positive (Beard et al., 2016; Heino

et al., 2018). The following models are estimated, respectively for the two dependent variables measuring trust in the media, resp. the scientific source:

$$MTrust_i = \beta_0 + \beta_1 T_i + \beta_2 X_{ik} + e_i \tag{4.1}$$

$$STrust_i = \beta_0 + \beta_1 T_i + \beta_2 X_{ik} + e_i \tag{4.2}$$

For each individual $i = 1, ..., n$, *MTrust* indicates the ordinal trust response in the media, *STrust* in the science source. T_i is an indicator of treatment assignment, whereas X_{ik} is a vector of k covariates, and e_i the error term.

Generally, the Bayesian approach allows for a straightforward interpretation of evidence for an hypothesis given the data $p(H|D)$, whereas the Frequentist approach estimates the probability of observing data given a hypothesis, usually the null hypothesis $p(D|H)$. In order to estimate $p(H|D)$ by applying *Bayes Theorem*, one needs to specify a prior for the investigated effect $p(H)$. A prior distribution reflects the subjective knowledge, uncertainty, or belief for the parameters before the researcher sees the data. Bayes Theorem then states:

$$p(H|D) = \frac{p(D|H) \times p(H)}{p(D)} \tag{4.3}$$

Here, the ratio on the left is the posterior probability, the left probability in the numerator the likelihood, which is multiplied with the prior, and the denominator is the marginal likelihood, which is a constant:

$$p(D) = p(D|H_0) \times p(H_0) + p(D|H_1) \times p(H_1) = const. \tag{4.4}$$

Consequently, expression (3) can be reduced to:

$$p(H|D) = p(D|H) \times p(H) \tag{4.5}$$

Based on this equation, it is possible to calculate the ratio, i.e. relative likelihood of two hypotheses.

$$\frac{p(H_1|D)}{p(H_0|D)} = \frac{p(D|H_1)}{p(D|H_0)} \times \frac{p(H_1)}{p(H_0)} \tag{4.6}$$

Here, the term on the left side is called the posterior odds, the term on the outer right the prior odds, and the middle term the Bayes factor (BF_{10}).[8] Thus, a Bayes factor is the weighed ratio of two likelihoods, i.e. a likelihood-ratio.

For point hypotheses the BF_{10} is the likelihood of data given the alternative hypothesis, divided by the likelihood of data given the null hypothesis (BF_{10}) (Heino et al., 2018). It is computed via the Savage-Dickey density ratio method, i.e. the ratio of the posterior density at the point of interest divided by the prior density at that point (see Wagenmakers et al., 2010). Then, the

[8] The index of the BF indicates which hypothesis is the nominator of the ratio: $BF_{01} = \frac{p(D|H0)}{p(D|H_1)}$

BF_{10} indicates to what degree the evidence for the point hypothesis has increased or decreased after seeing the data. For a directed hypothesis it is an *Evidence Ratio*, i.e. for an hypothesis test of $a > b$, it is the ratio of the posterior probability of $a > b$ and the posterior probability of $a < b$ (Bürkner, 2017). Intuitively, $BF_{10} = 10$ indicates that the data are ten times more likely under H_1 than under H_0. Such statements are not possible within the frequentist framework. Bayes factors have a range between 0 and infinity.

Per convention, Bayes factors between $\frac{1}{3}$ and 3 present "anecdotal" evidence. Substantial evidence is either indicated in favor of the null hypothesis when $BF_{10} < \frac{1}{3}$ or in favor of the alternative hypothesis when $BF_{10} > 3$. $BF_{10} = 1$ indicates equal support for both hypotheses. Such weighting of relative evidence is especially helpful for interpreting non-significant results, because, in a frequentist framework, non-rejection of the null hypothesis does not automatically allow for a scientific conclusion in its favor (Dienes, 2014; Lakens et al., 2018a).

In order to get a sense of how much the piror-specification influences the posterior distribution, and therefore the inferences made based on the Bayesian model, I take into account two alternative prior distributions. The first is a non-informative ("flat") prior drawn from a uniform distribution between -10 and 10. The boundaries are set in line with the reasoning of Gelman (2008), who argue that, in logistic regression, a change of 5 on the logistic scale is equivalent to a probability change from 0.01 to 0.5, or from 0.5 to 0.99. Consequently, changes of 10 on the logistic scale are rarely occurring because they would correspond to a change in probability of the outcome variable of 98 percentage points.

The second prior, a Cauchy distribution with center 0 and scale $\frac{1}{\sqrt{2}}$ is a default choice in psychological research (Morey et al., 2015). In comparison to the flat prior it expresses a prior belief in the absence of a treatment effect. Figure C.1 shows the two prior distributions in comparison with a Gaussian distribution with mean 0 and standard deviation 1. In all Bayesian regression models shown below, priors of the intercepts are drawn from a Cauchy distribution with location 0 and scale 10 (see Gelman, 2008).

Bayesian regression is conducted in R with the 'brms' package (Bürkner, 2017), using a cumulative normal response distribution with a logit link function with ordered thresholds (Kruschke, 2015). The model assumes that ordinal responses come from a latent metric variable, which is logistically distributed with mean μ and standard deviation σ. The thresholds θ_k on the continuous latent variable determine the corresponding ordinal responses. The probability of ordinal response k is given by:

$$p(y = k|\mu, \sigma, \theta_k) = \Lambda((\theta_k - \mu)/\sigma) - \Lambda((\theta_{k-1} - \mu)/\sigma) \tag{4.7}$$

where Λ is the standardized cumulative logistic function.

The 'brms' package uses the No-U-Turn Sampler (NUTS) (Bürkner, 2017). The regression provides a posterior distribution which depends on the chosen prior, which is updated by the data using Bayes' theorem (Heino et al., 2018). Tests for treatment effects on secondary outcomes, i.e. perceived

source expertise, trustworthiness, and content credibility, are conducted using linear regression.

The frequentist ordinal logit model is a cumulative logit model with proportional odds in which the dependent variable is the probability that it falls at or below a particular point. The model assumes that the treatment effect is identical for all response categories (Agresti, 2007).

4.4 Findings

4.4.1 Descriptive statistics

Table 4.1 shows subjects' perceptions of trustworthiness of the science, respectively media article source conditional on treatment assignment.[9] For both sources, aggregated as well as disaggregated by treatment the median perception of trustworthiness is 4. Summary statistics of the secondary outcome variables are in Table 4.2. Violin plots for primary and secondary outcome variables are in Appendix C.2.

TABLE 4.1: Trustworthiness by source and treatment

	Media trustworthiness					
	1	2	3	4	5	n
No Quote	1 (1.02 %)	8 (8.16 %)	14 (14.29 %)	45 (45.92 %)	30 (30.61 %)	98
Quote	0 (0 %)	4 (4.65 %)	20 (23.26 %)	41 (47.67 %)	21 (24.42 %)	86

	Science trustworthiness					
	1	2	3	4	5	n
No Quote	1 (1.04 %)	2 (2.08 %)	14 (14.58 %)	36 (37.5 %)	43 (44.79 %)	96
Quote	0 (0 %)	4 (4.6 %)	16 (18.39 %)	39 (44.83 %)	28 (32.18 %)	87

Notes: Shows the absolute and relative number of subjects per treatment group according to their reported media and science trustworthiness.

4.4.2 Treatment effects on primary outcomes

A Chi-squared test does not indicate that the distributions of media trustworthiness differs significantly between treatments ($\chi^2(4) = 4.403, p = 0.354$). Likewise, a one-sided Mann-Whitney test does not reject the hypothesis that media trustworthiness in the no quote condition is equal to or higher than

[9] For 10 subjects there is no data for media trust, while for 11 other subjects, there is no data for science trust. This is due to technical problems during the experiment. Available data of these 21 subjects is included in the respective analyses.

TABLE 4.2: Source and content credibility by source and treatment

	Media trustworthiness			Media expertise			Content credibility			
	Mean	SD	Median	Mean	SD	Median	Mean	SD	Median	n
No Quote	20.3	3.86	20	18.19	4.03	19	15.81	2.96	16	98
Quote	19.93	3.27	20	18	3.87	18	15.22	2.95	15	86

	Science trustworthiness			Science expertise			Content credibility			
	Mean	SD	Median	Mean	SD	Median	Mean	SD	Median	n
No Quote	21.64	3.55	22.5	22.1	3.58	23	17.46	3.04	18	96
Quote	20.78	3.37	20	21.82	3.19	22	17.2	2.57	18	87

Notes: Shows means, standard deviations, and medians for the constructed trustworthiness, expertise, and content credibility indexes by treatment.

in the quote condition in favor of the alternative, that trust in media is lower after perceiving doubt in media independence ($W = 4479.5, p = 0.215, d = 0.06$). An ordinal logit regression of media trustworthiness on treatment condition, as well as the covariates age, gender, study area, and contribution amount in the previous experiment is shown in Table 4.3. A table showing all estimators of covariates is in Appendix C.2. The *one-sided* test of the estimated treatment effect on media trustworthiness provides $p = 0.239$. Overall, there is no evidence in favor of H1.

F1: Increasing the salience of doubt in the independence of media does not reduce perceived trustworthiness of the source of the media article on climate change.

TABLE 4.3: Ordinal logistic regression of media and science trustworthiness on treatment and covariates, and treatment-order interaction

	Media trustworthiness			Science trustworthiness			Media trustworthiness			Science trustworthiness		
	B	CI_{95}	p	B	CI_{95}	p	B	CI_{95}	p	B	CI_{95}	p
Treatment												
Quote	-0.20	-0.76 0.35	0.478	-0.41	-0.96 0.15	0.152	-0.25	-1.06 0.55	0.536	0.18	-0.67 1.03	0.678
Media first							0.15	-0.65 0.94	0.717	1.18	0.37 1.99	0.004
Quote x Media first							0.12	-1.02 1.25	0.842	-1.02	-2.18 0.14	0.085
Covariates		Yes			Yes			Yes			Yes	
1|2	-6.63	-9.17 -4.10		-6.43	-9.10 -3.77		-6.02	-8.53 -3.51		-5.88	-8.50 -3.25	
2|3	-4.01	-5.71 -2.30		-4.44	-6.39 -2.48		-3.39	-5.05 -1.72		-3.87	-5.77 -1.96	
3|4	-2.46	-4.06 -0.86		-2.51	-4.32 -0.70		-1.85	-3.41 -0.28		-1.900	-3.66 -0.15	
4|5	-0.33	-1.88 1.23		-0.57	-2.33 1.19		0.29	-1.25 1.83		0.10	-1.63 1.83	
Observations		184			183			184			183	
Pseudo R^2		0.0287			0.035			0.030			0.055	
LR-chi^2		(df=10) 13.00			(df=10) 15.07			(df=12) 13.51			(df=12) 23.58	
AIC		467.5			441.106			471.0			436.6	

Notes: Shows point estimates (B), 95 %-confidence intervals (CI_{95}), and p-values (p) from the respective ordinal logistic regression model.

Table 4.4 shows findings from the Bayesian ordered logistic model. Using the flat prior provides a $BF_{10} = 20.74$, which suggests that the evidence in favor of no effect after seeing the data is 20.74 times the likelihood of no effect prior to seeing the data. For the directed hypothesis I observe $Evid.Ratio =$

3.3. This suggests that the alternative hypothesis, i.e. that reading the quote decreases trust in the news media, is 3.3 times more likely than the opposite treatment effect ($Estimate = -0.21, Est.Error = 0.29, CI_{95} = [-\infty; 0.28]$).[10] With the more informative Cauchy-prior I find that the evidence in favor of no effect after seeing the data is 2.98 times the likelihood of no effect before seeing the data. For the directed hypothesis analysis suggests that the hypothesis that reading the quote decreases trust in the news media is 2.81 times the likelihood of the opposite treatment effect ($Estimate = -0.17, Est.Error = 0.27, CI_{95} = [-\infty; 0.27]$). Neither the frequentist, nor the Bayesian hypothesis tests suggest that the data is unlikely under the null model, resp. that the alternative model is much more credible than the null model. However, the Bayes factors neither provide strong evidence in favor of the null model, as it would be the case if $BF_{10} < 1/3$.

TABLE 4.4: Bayesian logistic regression of media trustworthiness on treatment and covariates

	Media trustworthiness									
	Uniform(-10, 10)					Cauchy(0, $\frac{1}{\sqrt{2}}$)				
	Est.	Est. Err.	CI_{95}	ES	Rhat	Est.	Est. Err.	CI_{95}	ES	Rhat
Treatment										
Quote	-0.21	0.29	-0.78 0.37	3875	1.00	-0.17	0.27	-0.71 0.35	1706	1.00
Covariates			Yes					Yes		
Intercept 1	-7.13	1.4	-10.16 -4.66	2091	1.00	-7.19	1.42	-10.31 -4.73	1744	1.00
Intercept 2	-4.08	0.9	-5.82 -2.33	3286	1.00	-4.06	0.89	-5.79 -2.29	1686	1.00
Intercept 3	-2.49	0.85	-4.12 -0.85	3287	1.00	-2.46	0.83	-4.09 -0.83	1763	1.00
Intercept 4	-0.31	0.83	-1.95 1.33	3531	1.00	-0.28	0.81	-1.84 1.30	1911	1.00
Observations			184					184		
WAIC			463.06					467.4		

Notes: Shows posterior means, standard deviations (analogous to frequentist standard error), 95 % credible intervals, ES for effective sample size describing the number of efficient samples from the posterior distribution, and Rhat, which is the Rubin-Gelman convergence diagnostic. Samples were drawn with 4 chains, each with 2000 iterations, and 1000 warmup iterations, resulting in 4000 random draws from the posterior distribution. Samples were drawn using NUTS sampling. WAIC is the widely applicable information criterion. Findings are robust to the inclusion of covariates.

For H2, a Chi-squared test does not suggest that the distribution of science trustworthiness differs significantly between treatments ($\chi^2(4) = 4.658, p = 0.324$). A Mann-Whitney test rejects the hypothesis that science trustworthiness in the no quote condition is equal to or higher than in the quote condition in favor of the alternative, that trust in science is lower after perceiving doubt in media independence ($W = 4733, p = 0.048, d = 0.22$). An ordinal logit regression of science trustworthiness on treatment condition, as well as age, gender, study area, and contribution amount in the previous experiment is shown in Table 4.3. The *one-sided* test of the estimated treatment

[10] Note that the CI_{95} is the 95 % credible interval, which is the Bayesian version of the 95 % confidence interval. The former is a range of values on the posterior probability distribution that includes 95 % of the probability. After seeing the data, one believes with a certainty of 95 % that the true treatment effect is inside this interval, as opposed to a particular confidence interval, for which this probability is either 1 or 0 (Heino et al., 2018).

effect on science trustworthiness provides $p = 0.076$, which indicates non-significance.

Of the three frequentist tests, only the MWU is significant at 5 %, indicating that, given some assumptions on the underlying distributions, the medians are different. However, when controlling for covariates, the ordinal logit model does not indicate that reading the quote decreases trust in the science source.

F2: Increasing the salience of doubt in the independence of media does not reduce perceived trustworthiness of the source of the scientific article on climate change.

Table 4.5 shows the respective Bayesian models to gain insights in the strength of the evidence in favor of either hypothesis. The flat prior provides $BF_{10} = 10.54$, which suggests that the evidence in favor of no effect after seeing the data is 10.54 times the likelihood of no effect prior to seeing the data. Testing the directed hypothesis suggests that reading the quote decreases trust in the news media is 12.29 times more likely than the opposite effect ($Estimate = -0.41, Est.Error = 0.29, CI_{95} = [-\infty; 0.06]$). With the more informative prior, the evidence in favor of no effect after seeing the data is 1.69 times the likelihood of no effect before seeing the data. Testing the directed hypothesis suggests that reading the quote decreases trust in the science source is 9.44 times more likely than the opposite treatment effect ($Estimate = -0.34, Est.Error = 0.27, CI_{95} = [-\infty; 0.09]$). Although the frequentist tests do not unanimously reject the null-hypothesis, Bayesian analyses suggest that the evidence in favor of the directed alternative is stronger compared to the evidence in favor of the respective null hypothesis.

TABLE 4.5: Bayesian logistic regression of science trustworthiness on treatment and covariates

	Science trustworthiness									
	Uniform(-10, 10)					Cauchy(0, $\frac{1}{\sqrt{2}}$)				
	Est.	Est. Err.	CI_{95}	ES	Rhat	Est.	Est. Err.	CI_{95}	ES	Rhat
Treatment										
Quote	-0.41	0.29	-0.97 0.15	4000	1.000	-0.34	0.27	-0.88 0.16	2692	1.00
Covariates			Yes					Yes		
Intercept 1	-7.1	1.54	-10.55 -4.45	2395	1.000	-7.04	1.56	-10.64 -4.35	2105	1.00
Intercept 2	-4.58	1.02	-6.60 -2.64	4000	1.000	-4.52	1.03	-6.62 -2.56	2040	1.00
Intercept 3	-2.56	0.93	-4.41 -0.76	4000	1.000	-2.52	0.95	-4.46 -0.69	2041	1.00
Intercept 4	-0.58	0.91	-2.39 1.20	4000	1.000	-0.54	0.92	-2.39 1.26	2063	1.00
Observations			183					183		
WAIC			443.37					442.84		

Notes: Shows posterior means, standard deviations (analogous to frequentist standard error), 95 % credible intervals, the effective sample size describing the number of efficient samples from the posterior distribution, and Rhat, which is the Rubin-Gelman convergence diagnostic. Samples were drawn with 4 chains, each with 2000 iterations, and 1000 warmup iterations, resulting in 4000 random draws from the posterior distribution. Samples were drawn using NUTS sampling. WAIC is the widely applicable information criterion. Findings are robust to the inclusion of covariates.

Overall, while the data do not unanimously suggest that the quote negatively affects participants' trust in the media or science source, it is surprising that the quote appears to have a qualitatively stronger negative effect on trust in the science source, compared to the media source.

Further analyses reveal that subjects trust the science source more than the media source, aggregated over treatment condition ($V = 2256, p = 0.007$). While this is also true in the no-quote condition ($V = 814, p = 0.005$), this within-subject difference is not significant in the quote condition ($V = 366.5$, $p = 0.376$).[11] This fits the observation of a relatively and qualitatively stronger negative effect of the quote on science trust, compared to media trust. An ex-post rationalization of this finding is that the higher within-subjects trust in science (compared to media) allows for more wiggle-room to decrease this trust in light of the quote (between-subjects).

Furthermore, there are order effects. Trust in the science source is higher for subjects that read the media article prior to the science article, compared to those that read the science article first (aggregated over treatment conditions) ($W = 3292, p = 0.012$). This pattern is not apparent for trust in the media source ($W = 3901.5, p = 0.327$). Dis-aggregating by treatment condition reveals that the order effect for science trust is only occurring among subjects that do not read the quote (Table 4.3). One post-hoc interpretation of this pattern is that, if people read about climate change in the news media before they read an identical article in a scientific context, they tend to have higher trust in the scientific source afterwards. However, if distrust in the dependence of news media is salient, this relative trust-advantage of a scientific source appears to vanish. A possible explanation of this finding is discussed in the last section.

4.4.3 Treatment effects on secondary outcomes

Findings from linear OLS regression of factor-based scores for source trustworthiness, source expertise, and content credibility on treatment condition, including covariates, both for media and science source, are shown in Table 4.6 and Table 4.7, respectively.[12] As can be inferred from the tables, none of the treatment estimates are significantly different from zero at the conventional alpha level. Consequently, reading the quote does neither affect subjects' overall ratings of perceived source trustworthiness or expertise, nor the perceived content credibility of either the news media, or the scientific article. These findings are consistent with those presented in the previous section, corroborating the conclusion that there is no treatment effect. Tables with all estimators of covariates are in Appendix C.2.

[11] These explicit tests have not been pre-registered.

[12] Cronbach's alphas, as internal consistency measures of reliability of these tests are 0.91, 0.9, and 0.87 for media trustworthiness, expertise, and content credibility, respectively. For respective science tests, they are 0.92, 0.92, and 0.89, respectively.

TABLE 4.6: OLS-regression of media trustworthiness, expertise, and content credibility on treatment

	Media trustworthiness scale			Media expertise scale			Media content credibility		
	B	CI_{95}	p	B	CI_{95}	p	B	CI_{95}	p
Intercept	21.48	18.94 – 24.02	<.001	19.54	15.86 – 23.23	<.001	15.78	13.12 – 18.44	<.001
Treatment									
Quote	-0.33	-1.43 – 0.77	0.553	-0.19	-1.43 – 1.04	0.76	-0.52	-1.41 – 0.36	0.245
Covariates		Yes			Yes			Yes	
Observations		184			184			184	
R2 / adj. R2		.065/ .011			.051/ -.004			.070/ .017	
F-statistics		1.212			0.931			1.308	
AIC		1003.039			1040.645			931.385	

Notes: Shows point estimates (B), confidence intervals (CI_{95}), and p-values (p) from the respective OLS regression model. Robust confidence intervals.

TABLE 4.7: OLS-regression of science trustworthiness, expertise, and content credibility on treatment

	Science trustworthiness scale			Science expertise scale			Science content credibility		
	B	CI_{95}	p	B	CI_{95}	p	B	CI_{95}	p
Intercept	22.51	18.40 – 26.62	<.001	22.51	19.31 – 25.72	<.001	17.01	14.40 – 19.62	<.001
Treatment									
Quote	-0.71	-1.76 – 0.34	0.182	-0.27	-1.33 – 0.79	0.617	-0.22	-1.08 – 0.64	0.612
Covariates		Yes			Yes			Yes	
Observations		183			183			183	
R2 / adj. R2		.069/ .015			.043/ -.012			.036/ -.020	
F-statistics		1.269			0.782			0.648	
AIC		985.736			981.385			915.423	

Notes: Shows point estimates (B), confidence intervals (CI_{95}), and p-values (p) from the respective OLS regression model. Robust confidence intervals.

4.5 Discussion and Conclusion

This paper outlines an experimental study investigating the impact of proclaimed doubt in the independence of media on the perception of trustworthiness of a media and science article about climate change. Findings do not suggest that being confronted with such a proclamation of doubt negatively affects perceived trust in the source of either a media or scientific article on climate change. Therefore, the data do not indicate that proclamation of doubt leads to declining trust in media, and also does not appear to spill over to trust in science. Findings concerning the secondary outcome variables corroborate this interpretation.

Further analyses reveal two things: First, further research needs to check whether these are really null results, because Bayesian analyses do not unambiguously rate the null hypothesis as more credible than the alternative hypothesis, especially for H2. Second, although the overall trust in the scientific source is larger than in the media source, trust in the former relies on exposition to the media source prior to the scientific source. A post-hoc rationalization, and at the same time a caveat to inferences based on this experimental design, is that the subjects, i.e. German undergraduate students, were likely familiar with the news media source, but less likely so with the scientific source. If subjects read about the topic in a source they

know first, they also trust the less well-known one, compared to when they don't encounter the well-known source first. Furthermore, there is marginal evidence that this interacts with the treatment effect, which appears to exert a negative influence on science trust only when the media article was consumed first. This could be explained after the fact with a priming of distrust in less-familiar sources due to the quote. Further experiments need to control for the familiarity with different sources to get unconfounded, i.e. unbiased estimates of the treatment effect.

There is evidence that distrust in the media, or perceived media bias is especially prevalent among strongly conservative or Republican US citizens, especially when they are politically cynic (Lee, 2005). I believe it is fair to say that the sample of German, primarily undergraduate students does not correspond to these characteristics (Engels et al., 2013).[13] Because I expect that people with a prior believe in the systematic bias of media, as well as a profound distrust in the media, are most likely to respond to our treatment with a spillover of distrust in science, testing the treatment effect with German students is a very stringent and conservative test for this hypothesis.

There are some additional caveats to this experiment and some possible paths for future research. Because subjects were not informed that the quote was actually from a publicly and politically relevant person until after the experiment, further research should probe whether the hypothesized effects occur if subjects have such knowledge. Although thus, the experimental design at hand was the strictest test of the hypotheses, hypothesized effects could occur if doubt is proclaimed by a person with a central location in the recipient's social network. This is a promising starting point for future research that investigates the spillover potential of doubt proclamations conditional on the relative position of the sender in the social network of the receiver (see Turcotte et al., 2015). More precisely, interactions of treatment effects with party preferences of participants, or their political alignment with the source of the distrust-statement would be interesting to test. Additionally, in line with research on identity protective cognition (Kahan, 2010; Kahan, 2017), which describes a tendency of individuals to rate evidence conditional on the beliefs that predominate in their group, cultural similarities of consumers and sources of distrust-statements could be interesting research foci. This is especially the case since cultural similarities are fundamental components of echo chambers and filter bubbles, which are defined by the cohesion among their members with respect to beliefs, attitudes, and views on different matters, e.g. politics or climate change. Trust in news media and in the science on climate change strongly correlate with consumers' partisan predispositions and ideologies (Hmielowski et al., 2014; Ziegler, 2017) Additionally, future research could investigate effects of such statements in a more realistic environment. For example, the impact of similar statements via Twitter on consumers, or via Facebook could serve as interesting data sources to be

[13] Data on party preferences from another experiment (not yet published) based on the same population of potential participants shows the following distribution of party preferences: 21.07 % green, 18.87 % social democrats, 16.67 % christian democrats, 16.67 % left, 13.72 % non-voters, and 12.99 % others.

analyzed via means of text, and more specifically sentiment analysis (Bakshy et al., 2015; Quattrociocchi et al., 2016).

Altogether, findings presented here contribute to the emerging literature on fake news (Allcott and Gentzkow, 2017), which is currently fueled by its role in the 2016 US-presidential campaign, and due to the terms' contemporary loose usage by people in important political positions. Findings add to this literature insofar as they indicate to what extent people in such a position potentially affect public perceptions of media and science when expressing their personal, possibly motivated doubt in media independence. Specifically, evidence suggests no impact on the perception on the globally relevant issue of global man-made climate change of others facing such expressions. However, further research is needed to substantiate this conclusion.

5 Conclusion

Informed by behavioral environmental economics, this dissertation experimentally deals with two important aspects of a recent globally and socially relevant challenge: climate change. First, it investigates contemporary behavioral instruments that aim to motivate individual climate protection. Second, it investigates factors underlying the skepticism towards the news and science on climate change. Specifically, the thesis contributes to the extant literature by advancing the still relatively young science on pro-social nudges. It does so by answering questions on their effectiveness under specific conditions relative to conventional instruments, as well as by contributing to the still young research on factors underlying climate skepticism, especially prevalent in online environments.

The experimental studies presented here use advanced methodology to make causal claims, and to contribute to a positive long-term development of empirical science. This is not just important for the sake of science itself, but also for the credibility of social scientific findings and their perception in the public. It is crucial within the context of climate change science, which, in combination with public support, forms an important ground for policy-makers to create and advance climate protection policies.

Specifically, the three papers presented here entail the following most important conclusions: First, pro-environmental nudges can be efficient, even when they are implemented transparently by openly communicating their purpose and likely behavioral influence, even when people do not like their behavioral autonomy threatened. Second, conventional and behavioral instruments, i.e. defaults, recommendations, and mandatory minimum contributions to induce voluntary individual climate protection can be ineffective, especially when being not ambitious enough for highly intrinsically motivated individuals. While information on the source of the respective intervention may have an impact on their effectiveness, neither does source information cause motivation crowding, nor does it change the relative intervention effects. Third, while proclaimed doubt in the independence of news media does not appear to influence others in their trust-perceptions with respect to a news-, as well as scientific climate change article, further evidence needs to be collected in order to substantiate this conclusion.

There is still much research to be done on pro-social and pro-environmental nudges, and on factors influencing skepticism towards climate change. Specifically, while the effectiveness of defaults complemented by transparency information has been consolidated by several studies, further research could investigate whether this holds for all nudges, or only a specific sub-group. For example, social norms, or framings, might be affected by transparency, even if this is not the case for defaults. This is especially interesting since

the underlying working mechanisms, which can also overlap, differ between these instruments. Additionally, one could think of different types and intensities of transparency, potentially resulting in dissimilar impacts. With respect to the second paper discussed in this thesis, future research could investigate further into the optimal size of defaults relative to intrinsic motivation in pro-environmental decision contexts, since these appear to be the cause of unintended side-effects, such as motivation crowding-out. Also, changes in the underlying working mechanism of the default could lead to changes in their relative performance, as well as the potential interaction with source information. For example, source information might play a role if the interventions are interpreted by decision makers as implicit recommendations coming from interested, or (un-) informed parties (see Altmann et al., 2015). Following from the findings presented in the third paper, future research could take social networks, artificially created in the lab, or prior existing real-world networks in consideration as factors affecting how subjects respond to devaluations of the credibility of news outlets with respect to their attitudes towards news media and scientific articles on climate change. Especially partisan networks are of interest because of the likely dominant role of partisan beliefs on climate skepticism, particularly in the US (see Kahan, 2017).

In general, most of the findings from laboratory experiments, including those presented here, will profit from investigations in more realistic field experimental environments (Levitt and List, 2009). These can be artefactual field experiments using a more representative sample, or natural field experiments investigating a representative sample of decision makers in real decision situations with real consequences in natural decision environments (see Harrison and List, 2004). Realistic contexts are especially important for the investigation of nudges, which are commonly viewed as being effective because they rely on decision-making heuristics that developed during thousands of years of natural decision making. As such, nudges are most likely effective in contexts unlike artefactual laboratory experiments, where decision heuristics may be displaced by heuristics characteristic of laboratory contexts (Kraft-Todd et al., 2015).

Methodologically, future empirical and experimental research should take seriously the current developments following from the reproducibility-crisis in the field of (social) psychology and empirical economics. Preanalysis plans, a priori power analyses, as well as adequate statistical procedures allowing for a more thorough interpretation of findings should become a standard in the empirical and experimental sciences in order to increase their credibility and societal value. While this thesis strives to include such best practices, further improvements are necessary and possible. For example, the application of equivalence testing can be powerful, especially in combination with the application of Bayes factors, in order to reject the presence of a meaningful effect (instead of *no* effect) in hypothesis tests (Lakens et al., 2018a). Equivalence tests are extremely valuable complements for Bayes factors, since the latter's potential to quantify the evidence for the null is limited.

The Bayes factor always depends on the specification of the alternative hypothesis, and since there is an infinite number of alternative hypotheses it is logically flawed to infer from a specific Bayes factor that the null hypothesis is supported. The specific null and alternative hypotheses are just not all possible hypotheses that could be tested (Schimmack, 2018). These procedures, especially in combination with preregistered study designs, including a definition of the smallest effect size of interest, as well as an adequate a priori power analysis, have the potential to vastly improve the quality of empirical findings and inferences.

To conclude, finding solutions to climate change remains one of the biggest challenges for society on a global scale. However, the experimental papers presented in this thesis provide a humble effort to improve the understanding of behaviorally informed interventions that foster the voluntary individual mitigation of CO_2 emissions underlying climate change, as well as the current understanding of factors underlying the skepticism towards climate change. This is important for public acceptance of public policies, the underlying science, as well as science as a whole.

A

A.1 Experimental design

Instructions
Welcome and thank you very much for participating in this experiment. This experiment is about decision-making. Please read the following instructions carefully. Everything that you need to know in order to participate in this experiment is explained below. If you have any difficulties in understanding these instructions please raise your hand and I will come to you. Please note that communication between participants is strictly prohibited during the experiment. Communication between participants will lead to the exclusion from the experiment. The experimental procedure will be as follows. You will receive 10 Euro. Please decide how much of the 10 Euro you would like to spend on climate protection. You can choose freely how much, if any, you contribute to climate protection (whole numbers between 0-10). Should you decide to contribute, we will realize your contribution to climate protection by buying and retiring carbon emission licenses from the European Union Emissions Trading System (EU ETS) at the end of the experiment (please read the respective paragraph below for a description). By this, you have the possibility to make a real contribution to climate protection. The rest of the money is your private pay-out that you will receive in cash at the end of the experiment.

After making the decision you will be kindly asked to complete a short questionnaire. Please note that your decisions in this experiment are anonymous and will not be revealed at any stage to the other participants. (If relevant) a confirmation of the aggregated real payment to the climate protection fund will be sent to all participants at the end of the whole experiment.

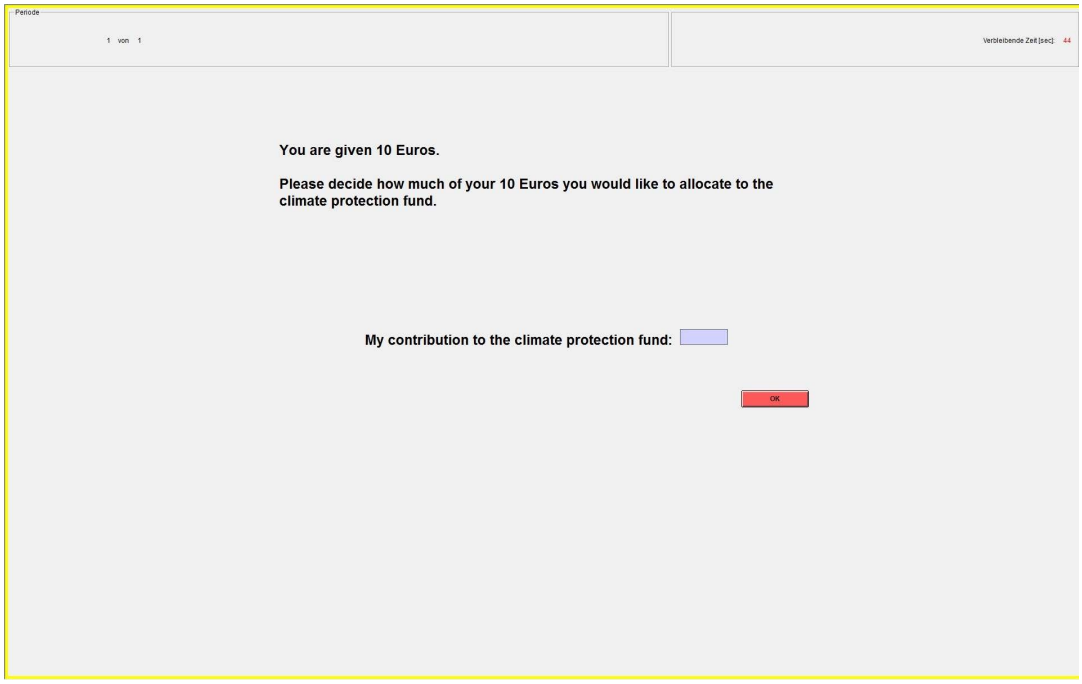
The Climate Protection Fund
If a person wants to protect the climate, emitting climate gases such as CO_2 should be avoided. But it is possible to do even more: Individuals can buy and delete emission certificates from the EU Emission Trading System (ETS) through certified organizations and NGOs. By doing so, a private person reduces the amount of CO_2 which can be emitted by European industries, protects the environment and ensures that the development of climate-friendly technologies is accelerated. In this experiment, the participants' contributions to the climate protection fund will be used to buy real carbon dioxide (CO_2) emission licenses on the market of the European Union Emissions Trading Scheme (EU ETS) via the website "TheCompensators.org". It is one example of an NGO that allows ordinary people to directly participate in the EU ETS scheme, and where they can make decisions on CO_2 reductions.

The following table shows how much kilograms of carbon you reduce with your payment, and how much money you receive for yourself. The far right row indicates the respective amount of reduced CO_2 relative to a Dutch citizens' average of 9163 kg of CO_2 emitted per year.

Payment to retire CO_2-allowances	**Private payout €**	**CO_2 abated [kg]**	**Share of average emissions per year per person [%]**
0	10 €	0	0%
1	9 €	181	2%
2	8 €	362	4%
3	7 €	542	6%
4	6 €	723	8%
5	5 €	904	10%
6	4 €	1,085	12%
7	3 €	1,266	14%
8	2 €	1,447	16%
9	1 €	1,627	18%
10	0 €	1,808	20%

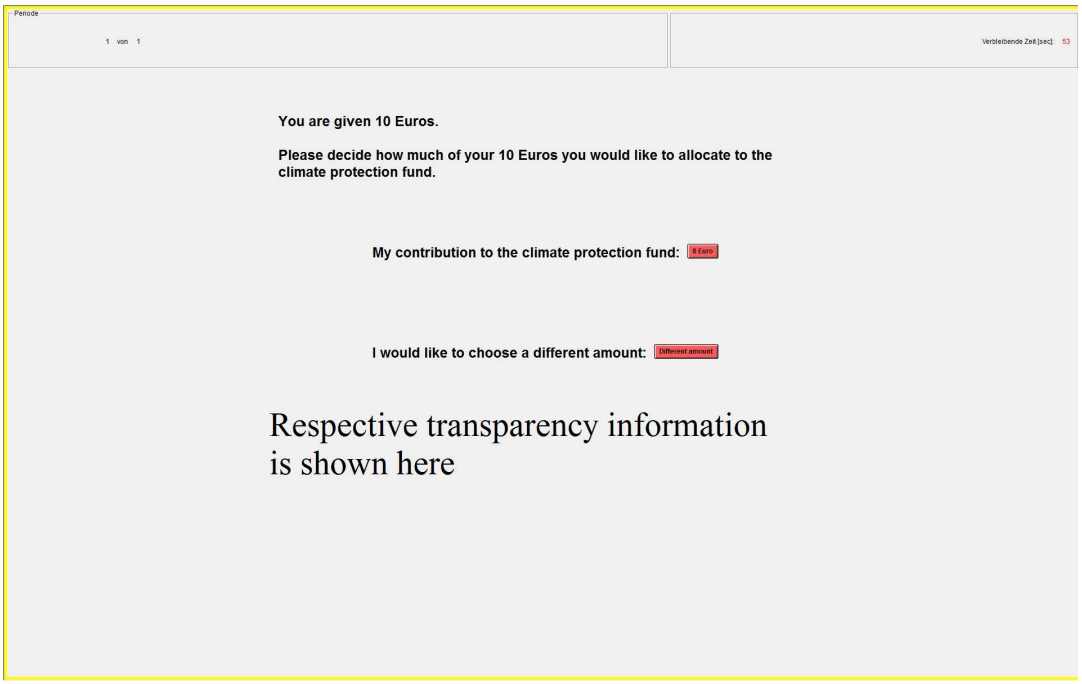
For example, with a payment of 3 Euro to retire carbon licenses, you retire 542 kg CO_2. This corresponds to approximately 6% of the average emissions per capita per year of a Dutch person. As a private pay-out you get 7 Euro. With a payment of 8 Euro to retire carbon licenses, you retire 1,447 kg CO_2. This corresponds to approximately 16% of the average emissions per capita per year of a Dutch person. As a private pay-out you get 2 Euro.

FIGURE A.1: Experimental screen for Control



Notes: The figure shows the decision screen shown to participants in the Control group. They could choose any integer between 0 and their endowment of 10 EUR. By clicking on the red OK button, subjects went to the next screen, providing them with information about the consequences of their decision, i.e. their payoff, their contribution, as well as kg of CO_2 offset.

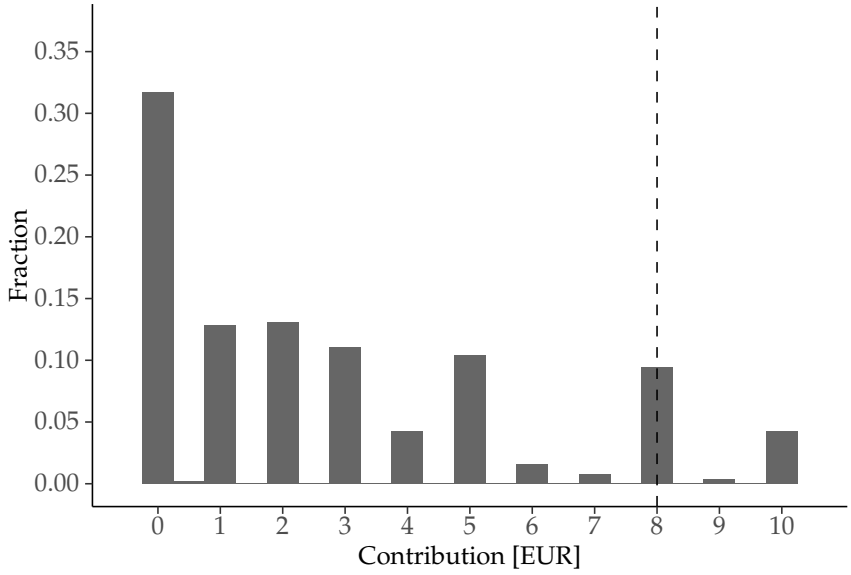
FIGURE A.2: Experimental screen for Default + transparency



Notes: The figure shows the decision screen shown to participants in the Default groups. They could choose to contribute the default value of 8 EUR by clocking on the respective red button, or they could click on the button below to choose any other amount. The transparency message was written where indicated in the figure. The following screen provided subjects with information about the consequences of their decision, i.e. their payoff, their contribution, as well as kg of CO_2 offset.

A.2 Statistical analyses

FIGURE A.3: Distribution of contributions



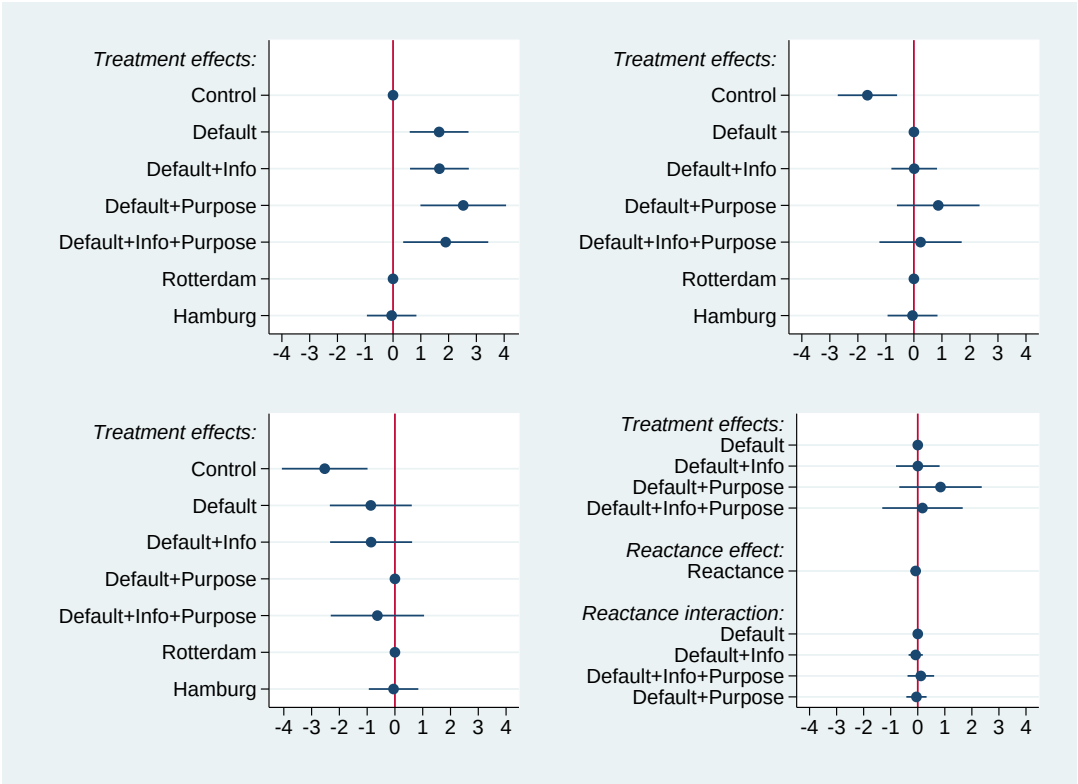
Notes: Shows the distribution of contribution amounts, more precisely the fraction of subjects contributing the respective amount. The dashed line indicates the default value.

TABLE A.1: P-values for pairwise MW tests of Contribution

	Control	Default	Default +Info	Default +Purpose
Default	**0.001**			
Default+Info	<**0.001**	0.665		
Default+Purpose	**0.032**	0.843	0.591	
Default+Info+Purpose	**0.046**	0.785	0.606	0.91

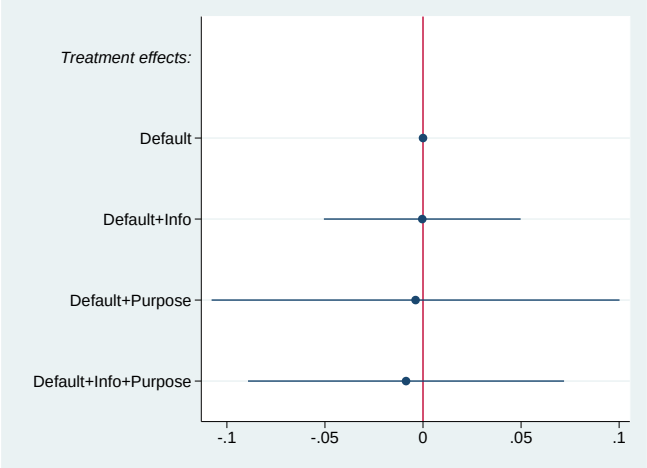
Notes: P-values of pairwise Mann-Whitney tests for equality of distributions of contributions to the climate protection fund. Comparisons are indicated by the treatment names provided in the first column and first row, respectively. Significance levels: $p < 0.05$ in bold, $p < 0.1$ in cursive.

FIGURE A.4: Default and transparency effects on contributions for different base-categories



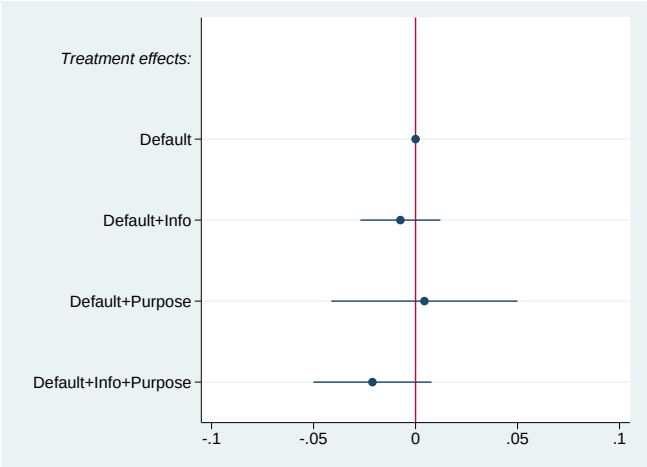
Notes: The figure graphically depicts results from some of the findings from the Tobit models. Dots with horizontal lines indicate point estimates with 95% confidence intervals. Dots on the zero line without confidence intervals denote the reference category. Models (3) and (8) in Table 2.4 display the underlying regression results. The top left panel refers to finding F1, the top right panel to F2 and F3, the bottom left panel to F4, and the panel on the bottom right to F6. Covariates are not shown.

FIGURE A.5: Default and transparency effects on perceived Threat to freedom



Notes: Dots with horizontal lines indicate point estimates with 95% confidence intervals from marginal effects of ordered logistic models. Dots on the zero line without confidence intervals denote the reference category. Model (4) in Table 2.5 displays the underlying regression results (albeit not showing marginal effects). It refers to finding F5. Covariates are not shown.

FIGURE A.6: Default and transparency effects on Anger



Notes: Dots with horizontal lines indicate point estimates with 95% confidence intervals from marginal effects of ordered logistic models. Dots on the zero line without confidence intervals denote the reference category. Model (5) in Table 2.5 displays the underlying regression results (albeit not showing marginal effects). It refers to finding F5. Covariates are not shown.

A.2.1 Comparing subjects from Rotterdam and Hamburg

We conducted experimental sessions in two different cities. Findings from the first eleven experimental sessions relied on data solely from Rotterdam, while additional observations where gathered in Hamburg primarily in order to increase the reliability of the null result presented in F2 (and to a minor degree F3-F4 by increasing the n in the control group). The number of additional observations gathered in Hamburg relied on an a priori power analysis. Based on this analysis we conducted additional sessions to gather 284 additional observations for the Control, Default, and Default+Info groups. The experimental protocol in all sessions was identical.

Table A.2 shows summary statistics of the main outcome variables disaggregated by treatment and location of the experiment. Contribution distributions in the Control ($W = 795.5, p = 0.329$), Default ($W = 3053.5, p = 0.528$), and Default+Info ($W = 2119.5, p = 0.092$) groups do not differ by location. The same is true for the remaining outcome variables. Figure A.7 shows the mean contributions disaggregated by location and treatments, including bars indicating 95 % confidence intervals. Mann-Whitney tests indicate that, while the default effect is significant in the Rotterdam sample ($W = 707.5, p = 0.007$), but not the Hamburg sample ($W = 2040.5, p = 0.074$), this is reversed with respect to the Default+Info effect, which is significant in Hamburg ($W = 1732.5, p = 0.009$), but not in Rotterdam ($W = 769.5, p = 0.084$). Differences between Default and Default+Info are insignificant in both samples (R: $W = 1113, p = 0.302$; H: $W = 6799, p = 0.24$)

Table A.3 shows summary statistics of the covariates included in the regression models disaggregated by treatment and location of the experiment. Aggregated over treatments, participants in Hamburg are on average older than participants in Rotterdam ($M = 24.94(SD = 4.81)$ vs. $M = 22.16(SD = 3.45), t(494.84) = -7.517, p < 0.001$), less likely to be male ($M = 39.08$ vs. $M = 57.01, \chi^2(1) = 15.038, p < 0.001$), and also have a different distribution of study areas ($\chi^2(6) = 156.65, p < 0.001$). Additionally, participants in Hamburg are more likely than their Rotterdam colleagues to rate climate protection as (very) important ($\chi^2(1) = 37.06, p < 0.001$). They do not differ with respect to prior experience in experiments ($chi^2(1) = 0.16, p = 0.69$) or their views regarding the effectiveness of the EU ETS ($\chi^2(1) = 0.002, p = 0.961$).

Aggregated over location, subjects are not balanced among treatments according to some variables. Subjects' ratings of the importance of climate protection correlate with the treatment ($\chi^2(4) = 34.37, p < 0.001$). So does age ($H(4) = 16.294, p = 0.003$), and the distribution of study areas ($\chi^2(6) = 156.65, p < 0.001$).

Figure A.8 shows standardized effect sizes and 95 % confidence intervals of the relevant pairwise comparisons for which we gathered additional data. While the effect size of the default effect (Con vs. Def) included zero in the Hamburg sample, it does not include zero in the Rotterdam- and the aggregate sample. The default+info effect size (Con vs. Def+Inf) is different from zero in the Hamburg and aggregated sample, but not in the Rotterdam sample. Although the standardized effect sizes for the Def vs. Def+Inf

comparison is opposite between Hamburg and Rotterdam, neither those nor the aggregated sample exclude an effect size of zero. Figure A.9 shows the regression coefficients and 95 % confidence intervals from Tobit model (3). These are qualitatively similar to the respective effect sizes, with the exception that the standardized effect size for the Con vs. Def comparison in Hamburg includes zero, whereas this is not the case for the respective regression coefficient.

TABLE A.2: Descriptive statistics of all outcome variables by experimental group and location

		Contribution		Contributed	Picked default	n
Group	Location	Mean	SD	Mean	Mean	
Control	R	1.67	2.68	46.67	0	45
Control	H	2	2.66	57.5	0	40
Default	R	3.24	3.21	73.91	19.57	46
Default	H	2.84	2.9	69.6	9.6	125
Default+Info	R	2.49	2.95	67.44	6.98	43
Default+Info	H	3.24	2.98	76.47	8.4	119
Default+Purpose	R	2.92	3.19	71.79	15.38	39
Default+Info+Purpose	R	2.85	2.95	65.85	17.07	41

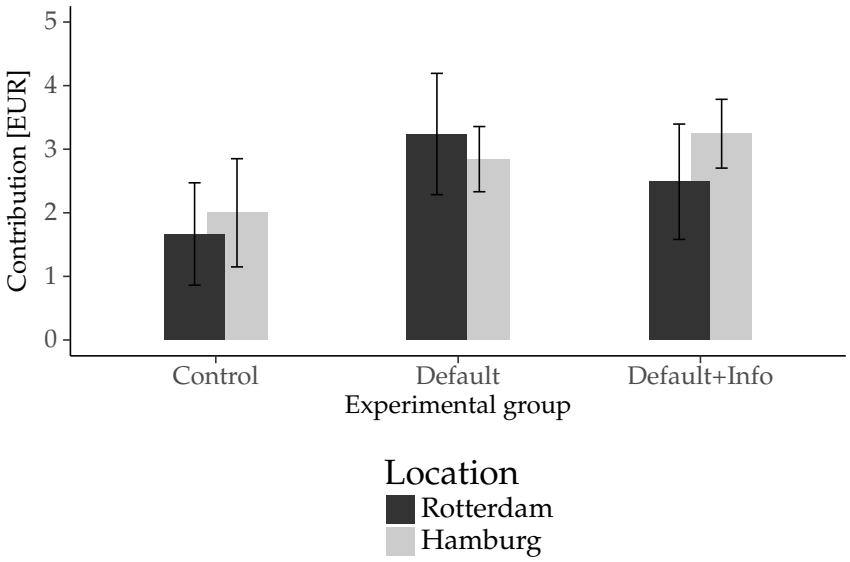
Notes: The table reports summary statistics (means and standard deviations) of different outcome variables, as well as the number of subjects per experimental group. Outcome variables are: contributions to the climate protection fund, the percentage of subjects contributing a positive amount, as well as the percentage of subjects contributing the default value. Statistics are disaggregated by experimental group and location of the experiment.

TABLE A.3: Descriptive statistics of covariates by experimental group and location

		Age		Gender (Male)	Importance of CP	No exp. Experience	EU ETS not effective
Group	Location	Mean	SD	Mean	Mean	Mean	Mean
Control	R	21.8	3.08	60	57.78	31.11	57.78
Control	H	25.95	5.7	35	97.5	15	62.5
Default	R	22.02	2.79	60.87	78.26	30.43	60.87
Default	H	24.95	4.48	36.8	84	29.6	60
Default+Info	R	22.07	2.96	51.16	79.07	20.93	53.49
Default+Info	H	24.59	4.81	42.86	91.6	27.73	57.98
Default+Purpose	R	22.28	4.65	53.85	51.28	20.51	64.1
Default+Info+Purpose	R	22.68	3.72	58.54	63.41	19.51	58.54

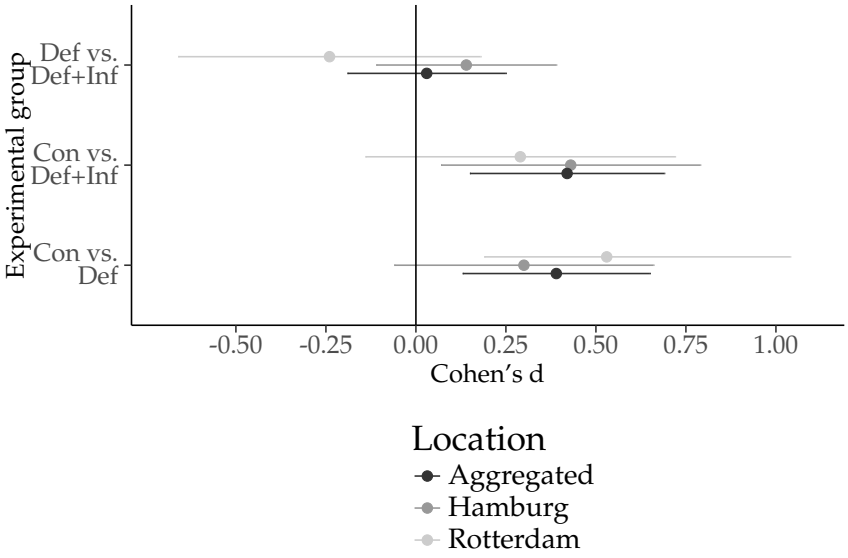
Notes: The table reports summary statistics (means and standard deviations) of different covariates per experimental group. Covariates are: age of participants, percentage of males, percentage of subjects perceiving climate protection as (very) important, percentage of subjects without prior experience with experiments, as well as the percentage of subjects judging license retirement as an ineffective mean for climate protection. Statistics are disaggregated by experimental group and location of the experiment.

FIGURE A.7: Mean contributions by experimental group and location



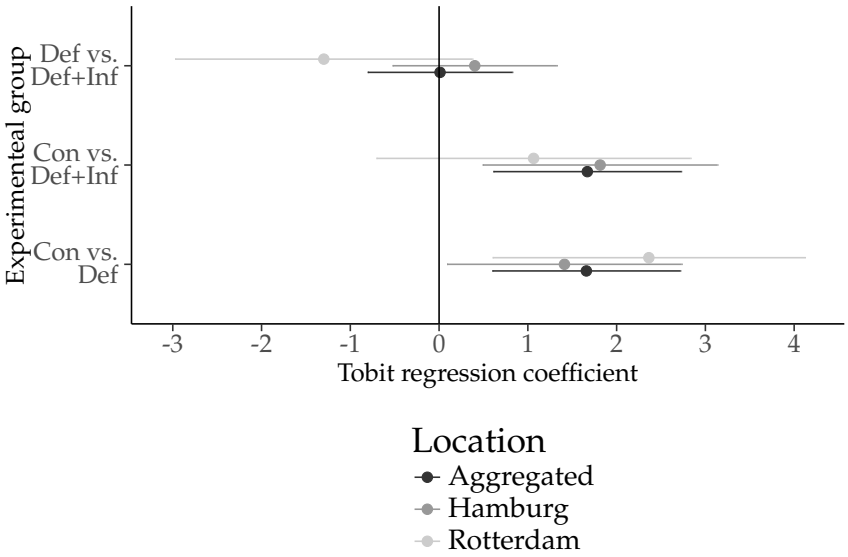
Notes: Shows mean contributions by experimental group and location, including 95 % confidence intervals.

FIGURE A.8: Effect sizes by location and for aggregated data



Notes: Shows Cohen's d for each pairwise comparison for which additional data in Hamburg was gathered, including the 95 % confidence intervals.

FIGURE A.9: Coefficients from tobit model by location and for aggregated data



Notes: Shows estimated coefficients from Tobit model (3) for effect for which additional data in Hamburg was gathered, including the 95 % confidence intervals.

A.3 Questionnaire

Questionnaire on covariates

What is you gender? O Male O Female

What is your age?

Have you participated in other experiments before today? O Yes O No

How important is climate protection for you? Please circle the most suitable answer.
O Not important at all O Not important O Indifferent O Important O Very important

Do you think that buying real carbon dioxide (CO_2) emissions licenses on the market of the European Union Emissions Trading Scheme (EU ETS) is an effective method to contribute to climate protection? O Yes O No

Questionnaire on state reactance

Please indicate to what extent do you agree with the following statements on a 5-point response scale that ranges from the statement – "strongly disagree" to the statement – "strongly agree". (Perceived threat to freedom)

- The default value threatened my freedom to choose.
- The default value tried to make a decision for me.
- The default value tried to manipulate me.
- The default value tried to pressure me.

Please indicate to what extent do you agree with the following statements on a 5-point response scale that ranges from the statement – "Not at all" to the statement – "Very". (anger)

- Please indicate how irritated you were with regard to the given default value.
- Please indicate how angry you were with regard to the given default value.
- Please indicate how annoyed you were with regard to the given default value.
- Please indicate how aggravated you were with regard to the given default value.

Questionnaire on trait reactance

Please indicate to what extent do you agree with the following statements on a p-point response scale that ranges from the statement – "strongly disagree" to the statement – "strongly agree".

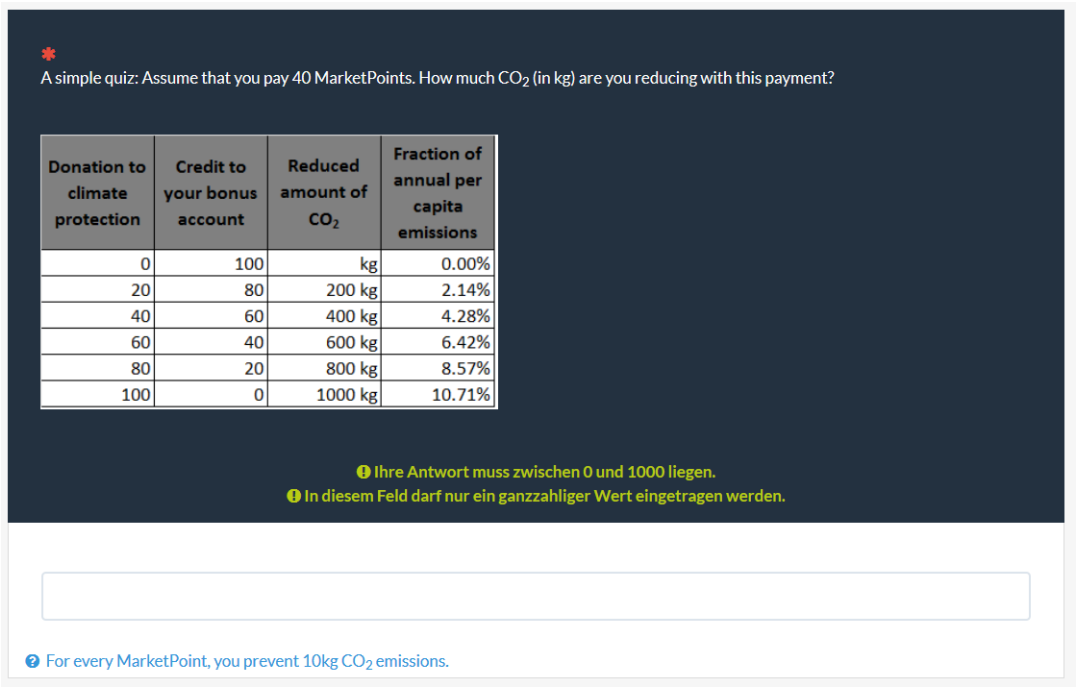
- Regulations trigger a sense of resistance in me.
- I find contradicting others stimulating.
- When something is prohibited, I usually think, "that's exactly what I am going to do".
- The thought of being dependent on others aggravates me.
- I consider advice from others to be an intrusion.
- I become frustrated when I am unable to make free and independent decisions.
- It irritates me when someone points out things, which are obvious to me.
- I become angry when my freedom of choice is restricted.
- Advice and recommendations usually induce me to do just the opposite.
- I am content only when I am acting on my own free will.
- I resist the attempts of others to influence me.
- It makes me angry when another person is held up as a role model for me to follow.
- When someone forces me to do something, I feel like doing the opposite.
- It disappoints me to see others submitting to standards and rules.

B

B.1 Experimental design

B.1.1 Experimental screens (translated)

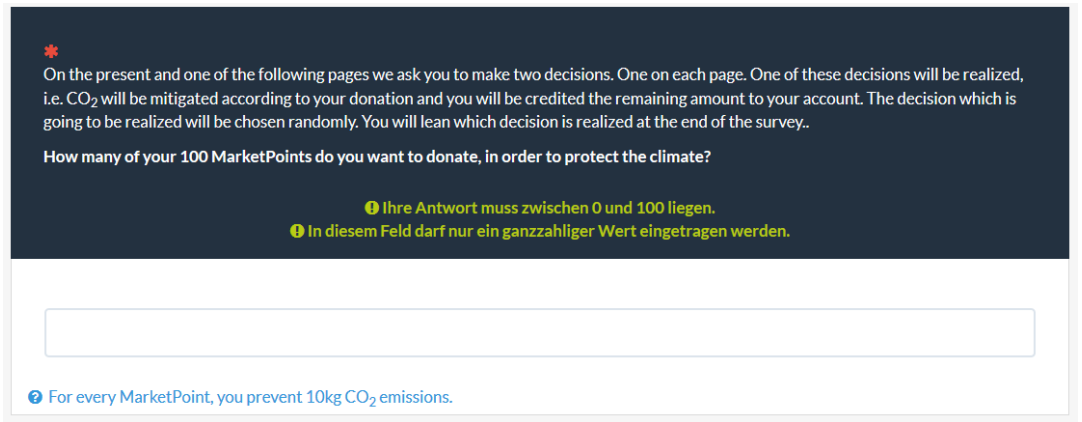
FIGURE B.1: Test question



Notes: The amount of MarketPoints (Credits) used in the quiz varies randomly between 10 and 90 Credits. Text in yellow reads "Your answer has to be in between 0 and 100." and "Only integers are allowed to be entered into this field."

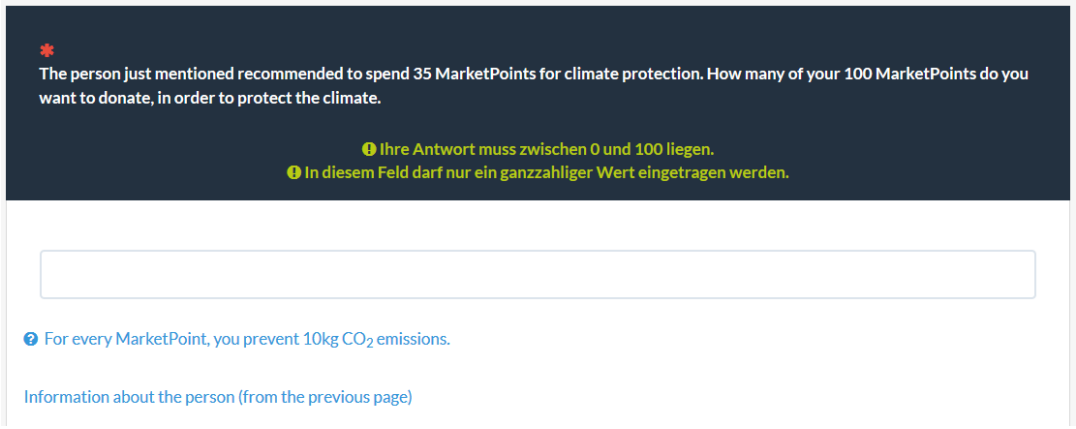
FIGURE B.2: Information about the source of the respective intervention (political)

FIGURE B.3: Decision screen for baseline contribution



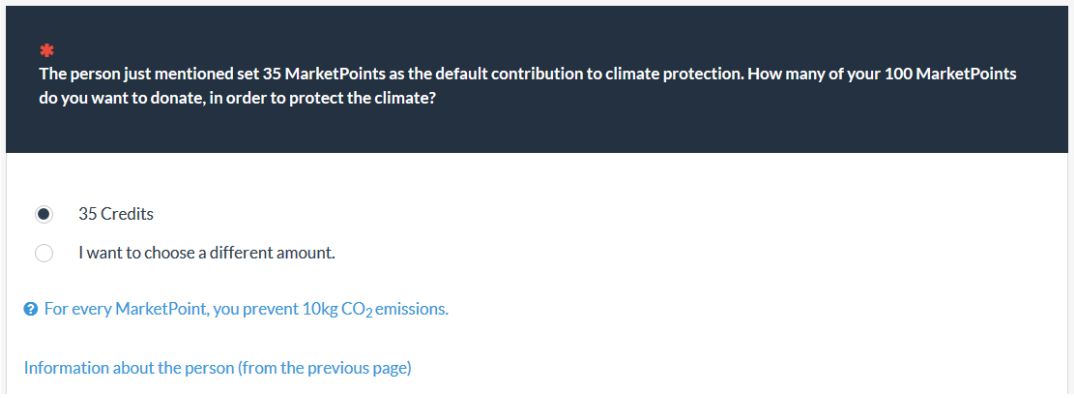
Notes: Text in yellow reads "Your answer has to be in between 0 and 100." and "Only integers are allowed to be entered into this field."

FIGURE B.4: Decision screen for recommendation intervention (with source information)



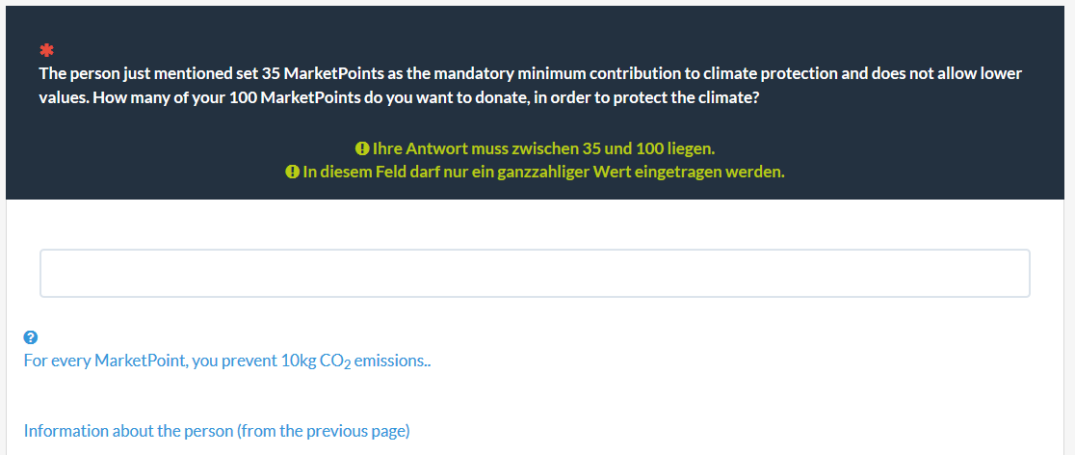
Notes: The respective text for the recommendation without source information reads "It is recommended to spend 35 Credits for climate protection". Subjects could click a button labeled "Information about the person (from the previous page)" not shown here. Text in yellow reads "Your answer has to be in between 0 and 100." and "Only integers are allowed to be entered into this field."

FIGURE B.5: Decision screen for default intervention (with source information)



Notes: The respective text for the default without source information reads "35 Credits were set as the default contribution to climate protection". Subjects could click a button labeled "Information about the person (from the previous page)" not shown here.

FIGURE B.6: Decision screen for restriction intervention (with source information)



Notes: The respective text for the restriction without source information reads "35 Credits were set as the mandatory minimum contribution to climate protection". Subjects could click a button labeled "Information about the person (from the previous page)" not shown here. Text in yellow reads "Your answer has to be in between 35 and 100." and "Only integers are allowed to be entered into this field."

B.1.2 Instructions (translated)

Welcome. Thank you very much for deciding to participate in this study. Please read the following instructions carefully. Everything you have to know will be explained on the following screens. Participation will take approximately 5-10 minutes. The course of this study will be as follows: You will receive 100 MarketPoints. During the course of this study you will be asked to indicate how much of these you want to donate to climate protection (you will receive a thorough explanation during the remainder of this study). Everything you do not donate will be added to your Lightspeed account. After stating your decision we ask you to will out a short questionnaire. The information you give in this study are anonymous. During data analysis we will not be able to assign your name to your answers and decisions. We assure you that this study and the data will be treated acknowledging data privacy. Data analysis will be done anonymously and findings will only be used for science. If you wish, you will receive certification of the overall reduced amount of carbon dioxide emissions made in this study within a month after the end of this study. Please note that the creator of this study committed to make only truthful statements.

How will climate protection be realized?

If you decide to donate money to climate protection in this study, we will buy the corresponding amount of emission rights and subsequently delete them. Thereby we reduce the European emissions of carbon dioxide (CO_2). That way you protect the climate. What you do in this study will thusly have a real effect on the climate. For every donated MarketPoint, 10kg CO_2 will be avoided.

The following table shows, by means of examples, how MarketPoints are translated to avoided CO_2 emissions, how many MarketPoints are transferred to your account, and how much CO_2 (in kilograms) you avoid with this. Die right column shows how many percent of yearly CO_2 emissions of an average German citizen you avoid with the respective donation.

An example: With a donation of *20 MarketPoints* you would reduce European emissions by 200 kg. This would translate to 2.14 % of yearly emissions of an average German citizen. You would take 80 MarketPoints for yourself.[1]

If you would like to know how exactly CO_2 emissions will be averted, please keep on reading. However, for the remainder of this study it is only important that your decisions are about real CO_2 emissions and real MarketPoints.

[1] The example donation is randomly drawn from the interval of integers $[1, 9] \times 10$ with the remaining values calculated accordingly.

Payment to retire CO_2-allowances	Private payout €	CO_2 abated [kg]	Share of average emissions per year per person [%]
0	10 €	0	0%
1	9 €	181	2%
2	8 €	362	4%
3	7 €	542	6%
4	6 €	723	8%
5	5 €	904	10%
6	4 €	1,085	12%
7	3 €	1,266	14%
8	2 €	1,447	16%
9	1 €	1,627	18%
10	0 €	1,808	20%

Information about the European emissions trading scheme and CO_2 emission rights

When burning natural gas, oil, gasoline, or coal to gain energy, carbon dioxide, so-called CO_2 emissions occur. CO_2 is regarded as one of the primary causes of man-made climate change. Hence, the European Union launched an emission trading scheme in 2005. It is based on a simple idea: a previously defined upper limit of CO_2 emissions for certain industry sectors is not allowed to be overstepped. These sectors consist of power production industries, industries with high energy consumption, as well as air traffic within Europe. The state hands out an appropriate amount of emission rights to the companies.

Every regulated company has give an emission right for each ton of emitted emissions to the state, without getting any money in return. Companies are allowed to trade emission rights with each other. Due to trade between companies a price for emission rights emerges. Consequently, emitting CO_2 costs money. If a company does not own any more emission rights, it is not allowed to emit any CO_2 any more. If it does nonetheless, it has to pay 100 € per ton CO_2 penalty.

As soon as the amount of emission rights gets reduced, the amount of CO_2 emissions allowed to be emitted also decreases. Buying emission rights and subsequently deleting them reduces CO_2 emissions that are overall emitted from the affected industries, with the same amount. Deletion is done by the NGO "TheCompensators".

B.2 Statistical analyses

TABLE B.1: Descriptive statistics of reactance proneness, environmental orientation, political orientation, and attitudes towards license retirement

	Trait Reactance score		NEP score		Party affiliation									Retirement not effective	n
					CDU	SPD	Green	Left	AfD	FDP	Pirate	NPD	n.a.		
	M	SD	M	SD	%	%	%	%	%	%	%	%	%	%	
Control	35.27	5.52	19.69	2.35	21.33	26.67	10.67	8.00	10.67	6.67	1.33	1.33	13.33	13.33	75
No Source															
Recommendation	35.33	6.54	19.01	2.60	31.11	14.44	1.11	14.44	11.11	5.56	5.56	1.11	15.56	13.33	90
Default	35.07	5.89	19.45	3.21	27.71	13.25	10.84	12.05	13.25	2.41	0.00	1.20	19.28	14.46	83
Restriction	34.83	5.58	19.36	2.37	29.07	17.44	13.95	5.81	13.95	2.33	2.33	0.00	15.12	12.79	86
Expert															
Recommendation	34.31	5.91	18.53	2.38	19.48	22.08	10.39	11.69	16.88	2.60	1.30	0.00	15.58	12.99	77
Default	35.08	6.21	19.60	2.21	27.40	12.33	8.22	12.33	16.44	2.74	2.74	1.37	16.44	10.96	73
Restriction	35.25	6.89	19.68	2.54	22.78	15.19	8.86	8.86	12.66	1.27	3.80	3.80	22.78	8.86	79
Politician															
Recommendation	36.06	5.75	19.59	2.58	18.07	19.28	16.87	13.25	14.46	0.00	1.20	0.00	16.87	12.05	83
Default	35.44	5.47	19.68	2.28	18.52	18.52	9.88	20.99	12.35	1.23	0.00	0.00	18.52	17.28	81
Restriction	35.54	7.04	19.23	2.15	24.05	17.72	3.80	10.13	17.72	6.33	0.00	1.27	18.99	18.99	79
Aggregated	35.22	6.08	19.38	2.50	24.07	17.62	9.43	11.79	13.90	3.10	1.86	0.99	17.25	13.52	806

Notes: Shows summary statistics for covariates and their distribution across treatments, as well as aggregated over treatments. M = mean, SD = standard deviation. Distributions of trait reactance do not vary between treatments ($H(9) = 4.18$, $p = 0.899$), neither does the NEP ($H(9) = 14.305$, $p = 0.112$), Party affiliation ($\chi^2(72) = 83.062$, $p = 0.175$), nor answers given to the question whether participants regard carbon retirement via the EUETS as ineffective ($\chi^2(9) = 5.155$, $p = 0.821$).

TABLE B.2: Descriptive statistics of gender, education, and income

	Age		Gender	Education					Income category											n
				In school	Low	Mid	High	Uni	< 600	600 -999	1,000 -1,249	1,250 -1,499	1,500 -1,999	2,000 -2,499	2,500 -2,999	3,000 -3,499	3,500 -3,999	4,000 -4,999	$> 5,000$	
	M	SD	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	
Control	49.75	16.04	45.33	0.00	34.67	33.33	17.33	14.67	11.59	21.74	15.94	17.39	26.09	21.74	23.19	14.49	11.59	15.94	15.94	75
No Source																				
Recommendation	48.60	16.51	50.00	0.00	32.22	35.56	12.22	20.00	19.23	19.23	19.23	25.64	30.77	28.21	28.21	30.77	21.79	23.08	23.08	90
Default	52.42	15.77	62.65	0.00	37.35	30.12	14.46	18.07	14.47	19.74	17.11	18.42	25.00	23.68	19.74	19.74	17.11	13.16	13.16	83
Restriction	49.17	16.76	52.33	0.00	36.05	32.56	17.44	13.95	16.46	15.19	16.46	20.25	22.78	27.85	26.58	15.19	12.66	11.39	12.66	86
Expert																				
Recommendation	52.26	14.93	42.86	0.00	25.97	41.56	15.58	16.88	11.43	18.57	24.29	14.29	21.43	20.00	27.14	15.71	20.00	18.57	18.57	77
Default	50.62	15.16	43.84	0.00	34.25	34.25	17.81	13.70	15.15	24.24	18.18	21.21	21.21	28.79	25.76	16.67	16.67	12.12	16.67	73
Restriction	49.54	14.65	56.96	1.27	37.97	29.11	11.39	20.25	23.08	32.31	26.15	29.23	33.85	43.08	41.54	33.85	26.15	23.08	24.62	79
Politician																				
Recommendation	49.06	15.00	48.19	0.00	31.33	26.51	16.87	25.30	17.33	20.00	21.33	14.67	30.67	24.00	22.67	14.67	17.33	16.00	18.67	83
Default	50.10	15.43	54.32	2.47	38.27	32.10	12.35	14.81	10.53	17.11	17.11	15.79	17.11	21.05	17.11	15.79	19.74	10.53	10.53	81
Restriction	47.72	16.23	60.76	0.00	29.11	39.24	12.66	18.99	19.12	29.41	29.41	29.41	33.82	25.00	23.53	27.94	23.53	20.59	16.18	79
Aggregated	49.90	15.66	51.86	0.38	33.75	33.37	14.76	17.74	15.79	21.47	20.36	20.50	26.18	26.18	25.35	20.36	18.56	16.34	16.90	806

Notes: Shows summary statistics for covariates and their distribution across treatments, as well as aggregated over treatments. M = mean, SD = standard deviation. Distributions of age do not vary between treatments ($H(9) = 6.727$, $p = 0.666$), neither does education ($\chi^2(36) = 30.11$, $p = 0.744$), nor income category of the household ($\chi^2(90) = 82.638$, $p = 0.697$).

TABLE B.3: P-values from pairwise Mann-Whitney tests for equality of contribution changes between experimental groups (excluding restriction treatments)

Experimental group	Experimental group					
	Control	Rec-No	Rec-Exp	Rec-Pol	Def-No	Def-Exp
Rec-No	0.092	-	-	-	-	-
Rec-Exp	0.389	0.014	-	-	-	-
Rec-Pol	0.883	0.048	0.421	-	-	-
Def-No	0.599	0.329	0.284	0.545	-	-
Def-Exp	0.668	0.233	0.231	0.548	0.873	-
Def-Pol	0.367	0.014	0.807	0.333	0.231	0.236

Notes: Shows p-values of pairwise Mann-Whitney tests with contribution changes as the dependent variable. Does not include treatments with a restriction on contribution levels. However, for these groups all pairwise comparisons to other interventions are significant at $p < 0.01$, while contributions among themselves are non-significant.

TABLE B.4: P-values from pairwise Mann-Whitney tests for equality of contribution changes > 35 between experimental groups

Experimental group	Experimental group								
	Control	Rec-No	Rec-Exp	Rec-Pol	Def-No	Def-Exp	Def-Pol	Res-No	Res-Exp
Rec-No	0.888	-	-	-	-	-	-	-	-
Rec-Exp	0.137	0.166	-	-	-	-	-	-	-
Rec-Pol	0.363	0.459	0.442	-	-	-	-	-	-
Def-No	0.016	0.022	0.274	0.055	-	-	-	-	-
Def-Exp	0.146	0.182	0.920	0.494	0.287	-	-	-	-
Def-Pol	0.008	0.011	0.221	0.034	0.958	0.223	-	-	-
Res-No	0.015	0.011	<.001	<.001	<.001	<.001	<.001	-	-
Res-Exp	0.244	0.199	0.007	0.025	0.001	0.009	<.001	0.156	-
Res-Pol	0.037	0.025	<.001	0.001	<.001	0.001	<.001	0.705	0.291

Notes: Shows p-values of pairwise Mann-Whitney tests with contribution changes > 35 as the dependent variable.

TABLE B.5: Linear regression models of intervention type and interaction with baseline contributions on contribution changes (> 35) without source information

	Contribution change >35			Contribution change		
	B	CI_{95}	p	B	CI_{95}	p
Intercept	0.86	-2.31 – 4.04	0.593	-0.44	-4.63 – 3.74	0.834
Intervention type						
Recommendation	0.39	-3.44 – 4.21	0.842	4.16	-1.12 – 9.45	0.122
Default	1.48	-2.51 – 5.47	0.466	3.31	-2.37 – 8.99	0.252
Restriction	3.65	-0.94 – 8.24	0.119			
BaseCon (a.)	-0.2	-0.41 – 0.01	0.063			
BaseCon (n.a.)				-0.22	-0.41 – -0.04	0.017
Interactions						
Rec. x BaseCon (a.)	-0.18	-0.50 – 0.15	0.287			
Def. x BaseCon (a.)	-0.76	-1.01 – -0.50	<.001			
Res. x BaseCon (a.)	-0.16	-0.49 – 0.17	0.344			
Rec. x BaseCon (n.a.)				-0.09	-0.32 – 0.15	0.469
Def. x BaseCon (n.a.)				-0.42	-0.66 – -0.19	<.001
Observations	334			248		
R^2 / adj. R^2	.391 / .378			.347 / .333		
F-statistics	29.893***			25.697***		
AIC	2717.228			2176.986		

Notes: Shows point estimates (B), 95 % confidence intervals (CI_{95}), and p-values (p) from robust OLS-regression. Control is the baseline for intervention type. Base contributions are median centered. (a.) means that baseline contributions are adjusted, i.e. that base contributions below 35 were set to 35. (n.a.) indicates no adjustment.

TABLE B.6: Linear regression models of source type and interaction with baseline contributions (> 35) on contribution changes (> 35), conditional on intervention types

	Recommendation						Default						Restriction		
	Contribution change >35			Contribution change			Contribution change >35			Contribution change			Contribution change >35		
	B	CI_{95}	p	B	CI_{95}	p	B	CI_{95}	p	B	CI_{95}	p	B	CI_{95}	p
Intercept	1.25	-0.88 – 3.38	0.249	3.72	0.50 – 6.94	0.024	2.34	-0.07 – 4.75	0.057	2.87	-0.97 – 6.71	0.143	4.51	1.20 – 7.82	0.008
Sourcetype															
Control	-0.39	-4.21 – 3.44	0.842	-4.16	-9.44 – 1.11	0.121	-1.48	-5.47 – 2.51	0.466	-3.31	-8.99 – 2.36	0.252	-3.65	-8.24 – 0.95	0.119
Expert	-0.74	-3.28 – 1.80	0.567	-5.07	-8.96 – -1.17	0.011	-1.19	-4.06 – 1.68	0.416	-1.68	-6.65 – 3.28	0.505	-1.46	-6.23 – 3.32	0.549
Politician	-1.71	-4.18 – 0.75	0.172	-4.99	-8.50 – -1.48	0.005	-2.13	-4.92 – 0.65	0.133	-3.21	-8.05 – 1.64	0.194	-1.99	-5.96 – 1.97	0.323
BaseCon (a.)	-0.38	-0.62 – -0.13	0.003				-0.96	-1.09 – -0.82	<.001				-0.36	-0.62 – -0.10	0.006
BaseCon (n.a.)				-0.31	-0.45 – -0.17	<.001				-0.65	-0.80 – -0.50	<.001			
Interactions															
Con. x BaseCon (a.)	0.18	-0.15 – 0.50	0.287				0.76	0.50 – 1.01	<.001				0.16	-0.17 – 0.49	0.344
Exp. x BaseCon (a.)	-0.13	-0.49 – 0.22	0.457				0.23	-0.08 – 0.54	0.14				0.08	-0.26 – 0.42	0.653
Pol. x BaseCon (a.)	0.11	-0.21 – 0.43	0.508				0.21	-0.08 – 0.51	0.157				0.1	-0.24 – 0.44	0.564
Con. x BaseCon (n.a.)				0.09	-0.15 – 0.32	0.469				0.42	0.19 – 0.66	<.001			
Exp. x BaseCon (n.a.)				-0.02	-0.23 – 0.18	0.837				0.15	-0.07 – 0.37	0.194			
Pol. x BaseCon (n.a.)				0.13	-0.05 – 0.31	0.144				0.11	-0.13 – 0.34	0.363			
Observations	325			325			312			312			319		
R^2 / adj. R^2	.311 / .296			.255 / .238			.612 / .603			.450 / .437			.169 / .150		
F-statistics	20.476***			15.465***			68.448***			35.483***			9.034***		
AIC	2571.043			2758.528			2413.36			2691.361			2592.237		

Notes: Shows point estimates (B), 95 % confidence intervals (CI_{95}), and p-values (p) from robust OLS-regression. No source is the baseline for source type. Base contributions are median centered. (a.) means that baseline contributions are adjusted, i.e. that base contributions below 35 were set to 35. (n.a.) indicates no adjustment.

TABLE B.7: Linear regression models of intervention type and source type, including interaction, on contribution changes (> 35)

	Contribution change >35			Contribution change		
	B	CI_{95}	p	B	CI_{95}	p
Intercept	-8.31	-12.85 – -3.77	<.001	-4.07	-9.85 – 1.71	0.167
Intervention type						
Recommendation	4.66	-0.95 – 10.26	0.103	4.22	-2.95 – 11.39	0.248
Restriction	8.09	2.15 – 14.04	0.008			
Source type						
Expert	0.09	-6.18 – 6.37	0.977	-0.64	-8.41 – 7.13	0.872
Polititian	-0.29	-6.18 – 5.60	0.923	-3.42	-11.00 – 4.15	0.375
Interactions						
Rec. x Exp.	-4.1	-12.26 – 4.06	0.324	-6.39	-16.24 – 3.47	0.203
Res. x Exp.	-0.99	-9.09 – 7.12	0.811			
Rec. x Pol.	0.03	-7.05 – 7.11	0.993	-0.2	-9.21 – 8.80	0.964
Res. x Pol.	-0.09	-7.58 – 7.39	0.98			
Observations		731			487	
R^2 / adj. R^2		.040 / .029			.015 / .005	
F-statistics		3.736***			1.467	
AIC		6181.43			4348.455	

Notes: Shows point estimates (B), 95 % confidence intervals (CI_{95}), and p-values (p) from robust OLS-regression. Default is the baseline for intervention type. No source is the baseline for source type.

TABLE B.8: Linear regression model for carryover effects

	Contribution		
	B	CI_{95}	p
Intercept	27.07	20.72 – 33.42	<.001
Two	8.23	-1.97 – 18.42	0.113
Default	-0.63	-8.64 – 7.39	0.878
Two x Default	-7.61	-20.05 – 4.83	0.229
Observations		313	
R^2 / adj. R^2		.017 / .007	
F-statistics		1.745	
AIC		2977.866	

Notes: Shows point estimates (B), 95 % confidence intervals (CI_{95}), and p-values (p) from robust OLS-regression. *Two* is a dummy taking the value 1 if there was a baseline decision, 0 otherwise. *Default* takes the value 1 if the treatment was a default, 0 otherwise.

B.2.1 Information on participant selection

Overall, 2,305 panelists from a German panel company were invited to the first stage of the experiment. They had participated in online surveys more or less frequently before, and were thus accustomed to similar types of surveys. Between November 29th, 2016 and December 7th, 2016, 1,436 panelists filled out the questionnaire (Stage 1). The other 869 panelists did not complete the questionnaire, which can be for two reasons: First, a subject can simply decide to not finish, or forget to finish the questionnaire, at any point. Second, a subject can be screened out based on its gender, education, and/or age, depending on whether the respective quota had already been reached. Quotas, i.e. upper limits for the number of observations satisfying certain criteria with respect to age, gender, and education level were implemented in order to achieve a sample representative of the German internet using population. If subjects had a combination of age, gender, and education that had already the necessary number of observations, they were "screened-out" of the experiment, meaning that they could not participate. Although ex-post we cannot infer the reason for not participating for each subject, we can definitely exclude the second explanation for the 37 participants that do not drop out directly after the socio-demographic questions. The remaining 832 drop-outs occurred either due to reached quotas, or due to voluntary drop-outs. However, we know by monitoring the experiment that the majority was screened-out shortly before Stage 1 ended, while "filling" the last quotas. Of all subjects in Stage 1 that did not choose to drop out and that were not screened out, 19 subjects completed the questionnaire in less than 20 seconds. These subjects were not expected to participate attentively and were therefore excluded.

Of the participants that filled out the questionnaire in Stage 1, 472 were later randomly assigned to Block 1, and 945 were assigned to Block 2 of Stage 2. Block 1 included two experimental groups (Control and Def-Exp) each in two versions (once with and once without baseline contribution decision). This part of the experiment was used to decide on an experimental parameter (assessing baseline contributions, or not) for the actual experiment (see section 3.4 for more detail). Block 2 included the remaining treatments that included the baseline contribution decision. By means of permuted block randomization, the four treatments in Block 1 were assigned to subjects, and the eight treatments in Block 2 were assigned to the remaining subjects.

Between December 14th, 2016 and January 2nd, 2017, 365 panelists participated in Block 1. Reminders to participate for all those that had not yet started or completed were sent per mail at the 19th, 27th, and 30th of December. Of all invited to this block, 107 (22.67 %) never started the survey. Of all that started the survey, 48 (13.15 %) did not finish.[2] Of all subjects that did not choose to drop out in the first block, 2 subjects were labeled as speedsters, meaning that they completed the questionnaire in less than 40, resp. 50

[2] Overall, 37 (77.12 %) of those that did not finish quit before or on the screen showing the test question, i.e. after or on the screen explaining the experiment and how carbon dioxide emissions were going to be retired, but before the (first) contribution decision.

seconds, depending on the treatment.

Between January 5th, and January 19th, 2017, 770 panelists participated in the second block of Stage 2. Reminders were sent to all those that had not yet started or completed the questionnaire at the 10th, 13th, and 16th of January. Of all invited to this block, 175 (18.52 %) never started the survey. Of all that started the survey, 105 (13.64 %) did not finish. There were no respondents that finished too fast.

Overall, of the 1,417 participants that were invited to Stage 2, 982 (69.30 %) completed the experiment (317 (67.16 %) in Block 1 and 665 (70.37 %) in Block 2). This amounts to an overall attrition rate between both stages of 31.70 % (32.84 % in Block 1 and 29.63 % in Block 2). Note that attrition is the sum of subjects that drop out and that do not start the survey, divided by the number of invited subjects.

We tested if subjects that dropped out of the experiment significantly differ from subjects that do not. We have socio demographic data on 143 subjects that dropped out. On average, drop-outs are older (50.13 vs. 54.78; $t(183.22) = -3.275, p < 0.01$), but do not differ with respect to the distribution of gender ($\chi^2(1) = 1.320, p = 0.251$), education (FE test: $p = 0.873$), or income category ($\chi^2(10) = 12.082, p = 0.280$).

We cleaned the data from non-completes, speedsters, as well as from participants that indicated they knew the source of the intervention, i.e. the German politician, from before the experiment.[3] The latter was the case for 9 subjects. Additionally, we did not include participants from the first block that faced only one contribution decision (165 observations). Consequently, final data consist of 806 full observations, for which we are able to match measures from both stages of the two-round version of the experiment.[4] Table B.2 shows the distributions of these covariates disaggregated by experimental group and aggregated (last row). Table B.1 shows the distributions of answers given to central questions from the questionnaire. On average, the median participant took 3 minutes in the first stage to finish the questionnaire. In Stage 2, the median participant took 7 minutes.

[3] The reason is that in this case we lose control over the source-manipulation, because subjects may be influenced by other factors of the source than those we elicit.

[4] In our pre-analysis plan, we planned to invite 1,370 subjects. Assuming an attrition rate of 30 %, we expected to have 960 final observations.

B.3 Questionnaire

B.3.1 Pre-experimental questionnaire (Stage 1)

Trait reactance

Please indicate to what extent you agree with the following statements. (Not agree at all, Not agree, Undecided, Agree, Strongly agree)
Order was randomized

- I find contradicting others stimulating.
- It makes me angry when another person is held up as a role model for me to follow.
- Regulations trigger a sense of resistance in me.
- When something is prohibited, I usually think, "that's exactly what I am going to do".
- I consider advice from others to be an intrusion.
- I become frustrated when I am unable to make free and independent decisions.
- It irritates me when someone points out things, which are obvious to me.
- I become angry when my freedom of choice is restricted.
- Advice and recommendations usually induce me to do just the opposite.
- I resist the attempts of others to influence me.
- When someone forces me to do something, I feel like doing the opposite.

New ecological paradigm (NEP)
Please indicate to what extent you agree with the following statements. (Not agree at all, Not agree, Undecided, Agree, Strongly agree)
Order was randomized

- The balance of nature is strong enough to cope with the impacts of modern industrial nations.
- Humans have the right to modify the natural environment to suit their needs.
- Humans are severely abusing the environment.

- Humans were meant to rule over the rest of nature.
- The balance of nature is very delicate and easily upset.
- Plants and animals have as much right as humans to exist.

Political orientation
Irrespective of how you expect the upcoming German federal election 2017 to end we would like to know, what you wish will happen. What would you prefer, personally: Which party shall get the most votes?

- CDU/CSU
- SPD
- FDP
- Die Linke
- Bündnis 90/Grüne
- Piratenpartei
- AfD
- NPD
- I don't want to say
- Other party

B.3.2 Post-experimental questionnaire (Stage 2)

Beliefs
How much do you think other participants contribute on average to climate protection?

Actor specific questions *Only the second question was asked to all subjects, whereas the remaining questions were only shown to subjects in source-treatments.*

- Did you know the previously mentioned person before participating in this experiment? (Yes/ No)
  Possible answers for following questions: Not agree at all, Not agree, Undecided, Agree, Strongly agree
- Reducing carbon emission rights is an effective way to reduce carbon emissions.
- The previously mentioned person is knowledgeable and competent with respect to climate protection.
- The previously mentioned person tries to influence me in my free decision.
- The previously mentioned person tries to reduce carbon emissions.

State reactance
Please indicate to what extent you agree with the following statements.
(Not agree at all, Not agree, Undecided, Agree, Strongly agree)

Threat to freedom

- The *intervention type* threatened my freedom to choose.
- The *intervention type* tries to make a decision for me.
- The *intervention type* tries to manipulate me.
- The *intervention type* tries to pressure me.

Anger

- The *intervention type* irritated me.
- The *intervention type* made me angry.
- The *intervention type* annoyed me.
- The *intervention type* aggravated me.

C

C.1 Instructions and materials

Instructions

German (original): Wir bitten Sie im Verlauf dieses Experiments 2 bis 3 Texte zu lesen. Bitte lesen Sie die Texte ausführlich und in Ruhe durch. Lassen Sie sich dafür ausreichend Zeit. Nach jedem Text bitten wir Sie darum, uns Ihre Eindrücke zum **Inhalt** und **Autor** von zwei Texten anzugeben. Sie werden dazu Ihre Zustimmung zu bestimmten, den **Textinhalt** und den **Autor** beschreibenden Adjektiven angeben. Sollten Sie Fragen haben, machen Sie bitte mit einem Handzeichen auf sich aufmerksam. Achten Sie bitte darauf, dabei andere Teilnehmer nicht zu stören. Wenn Sie die Anleitung verstanden haben, klicken Sie bitte auf OK und beginnen Sie mit dem Experiment.

English (translated): For the following experiment we ask you to read two to three texts. Please read all of them thoroughly and carefully. Give yourself enough time for that. After each text we kindly ask you to tell us your impressions regarding the **content** and **source** of two texts. You will express your level of agreement to different adjectives describing the **content** and **source** of the article. If you have a question please raise your hand and someone will assist you. Please make sure that you do not disturb other participants. IF you understood these instructions please click on OK and start with the experiment.

Media article
German (original): Dürren, Zyklone, schrumpfende Gletscher
2015 war das wärmste Jahr seit Beginn der Messungen – mit gravierenden Folgen. Diese gehen laut US-Klimabehörde auf den Klimawandel und das Wetterphänomen El Nino zurück. Die Rekordtemperaturen des vergangenen Jahres haben an vielen Orten der Welt gravierende Folgen nach sich gezogen. Wie die US-Klimabehörde NOAA mitteilte, zählten dazu Dürren, schrumpfende Gletscher, Fischwanderungen und Zyklone. Die Durchschnittstemperatur über Landflächen habe den Rekord von 2014 um mehr als 0,1 Grad Celsius übertroffen. Dazu habe auch das Klimaphänomen El Niño beigetragen, das aufgrund von veränderten Luft- und Meeresströmungen weltweit Wetterbedingungen verändert. Demnach gab es im vergangen Jahr auch die bislang höchste gemessene Konzentration von Treibhausgasen in der Luft. So überschritt laut NOAA die Durchschnittskonzentration von Kohlendioxid an

der Messwarte am Mauna Loa auf Hawaii erstmals 400 ppm (Teile pro Million Teile). Auch die Durchschnittstemperatur der Meere war 2015 so hoch wie nie zuvor seit Messbeginn.

NOAA-Direktor Thomas Karl sagte, der jährliche "ärztliche Check-up"der Erde zeige, dass der langfristige Klimawandel und El Nino das vergangene Jahr geprägt hätten. Zu den Symptomen zählt die NOAA eine riesige schädliche Algenblüte im nordöstlichen Pazifik mit deutlichen Auswirkungen auf die anderen Meereslebewesen und die davon abhängenden Menschen. Zudem seien die Berggletscher nach vorläufigen Daten weiter geschrumpft – das 36. Jahr in Folge. Die Zahl der tropischen Zyklone nahm deutlich zu: Mit insgesamt 101 Zyklonen lag sie deutlich über dem Schnitt, der zwischen 1981 bis 2010 noch 82 betrug.

Auch die Tiere der Arktis seien von den Entwicklungen beeinflusst worden, heißt es in dem Bericht. So zögen sich Walrossherden beispielsweise auf das Land zurück, anstelle auf Eis über dem Meer zu bleiben. In der Barentssee nördlich von Norwegen zögen die steigenden Temperaturen Warmwasserfische an und vertrieben andere einheimische Arten. Für den jährlichen Bericht werteten mehr als 450 Wissenschaftler aus rund 60 Ländern Daten aus.

English (translated): Droughts, Cyclones, shrinking glaciers.
2015 was the warmest year since the beginning of measurements - with grave consequences. Based on a US-climate agency these consequences are caused by global warming and the weather phenomenon El Nino.
Record temperatures in the last year had grave consequences on many places in the world. The US-climate agency NOAA reported that droughts, shrinking glaciers, fish migration, and cyclones were part of this development. Average temperatures above land broke the record of 2014 by more than 0.1 degrees Celsius. The weather phenomenon El Nino, affected by the changing air- and sea-currents, contributed to that.

According to that, the last had the highest concentration of greenhouse gases measured so far in the air. NOAA reported that the average concentration of carbon dioxide measured at the control room in Mauna Loa, Hawaii exceeded 400 ppm (parts per million) for the first time. The average temperature of the sea was as high as never before in 2015, as well.

The director of the NOAA, Thomas Karl, said the annual "check-up" for the plant shows that long-term climate change and El Nino shaped the preceding year. The NOAA counts a giant algae bloom in the northeastern Pacific Ocean with its obvious consequences for other marine organisms and people depending on them to these symptoms. Moreover, the mountain glaciers, based on preliminary data, proceeded to shrink for the 36th consecutive year. The number of tropical cyclones increased considerably: With 101 cyclones it was significantly higher than the average of 82 that was measured between 1981 to 2010.

Also the animals in the Arctic were supposedly influenced by these developments, writes the report. For example, walrus flocks retreated to the

land instead of remaining on the sea ice. In the Barents Sea north of Norway increasing temperatures supposedly attracted warm water fishes and displaced domestic species. For the annual report more than 450 scientists from roughly 60 countries analyzed data.

Science article
German (translated): Internationaler Report bestätigt, dass die Erde heiß ist und heißer wird.
2015 löst 2014 als wärmstes Jahr seit Beginn der Messungen ab - mit Hilfe von El Nino.
Ein jährlicher Klimabericht hat bestätigt, dass 2015 das Jahr 2014 als wärmstes Jahr seit mindestens Mitte des 19. Jahrhunderts ablöst. Laut über 450 zu dem Bericht beitragenden Wissenschaftlern resultiere die Rekordhitze des vergangenen Jahres aus einer Kombination des langfristigen Klimawandels und des stärksten seit 1950 gemessenen El Nino. Sie fanden außerdem, dass die meisten Indikatoren des Klimawandels weiterhin den Trends einer globalen Erwärmung entsprächen.

Weitere relevante Erkenntnisse:

- Es wurden die bisher höchsten Konzentrationen von Treibhausgasen gemessen. Die bedeutendsten Treibhausgase, Kohlenstoffdioxid (CO_2), Methan und Distickstoffoxid stiegen 2015 auf neue Rekordwerte. Die globale CO_2 Konzentration in diesem Jahr betrug 399.4 ppm (Teile pro Million Teile), ein Anstieg um 2,2 ppm im Vergleich zu 2014.

- Die globale Oberflächentemperatur war die höchste seit Beginn der Aufzeichnung. Begünstigt durch den starken El Nino betrug die jährliche globale Oberflächentemperatur $0,42\,^{\circ}C$ - $0,46\,^{\circ}C$ mehr als der Durchschnitt zwischen 1981 - 2010 und löste damit den bisherigen Rekord von 2014 ab.

- Die Durchschnittstemperatur der Meere war die höchste seit Messbeginn. Die globale Durchschnittstemperatur der Wasseroberfläche war $0,33\,^{\circ}C$ - $0,39\,^{\circ}C$ höher und löste damit den bisherigen Rekord von 2014 ab.

- Der Wärmeinhalt der oberen Ozeanschichten war ebenso auf Rekordhöhe und spiegelte die ansteigende Konzentration von Wärme in den oberen Ozeanschichten wieder.

- Der weltweite Wasserspiegel erreichte 2015 einen neuen Rekord. Mit 70mm höheren Werten als in 1993, als die Aufzeichnung durch Satelliten der globalen Wasserspiegelanstiege begann.

- Die Anzahl tropischer Zyklone lag insgesamt deutlich über dem Schnitt. Mit 101 tropischen Zyklonen über allen Ozeanen in 2015 lag die Anzahl

deutlich über dem Schnitt von 82 zwischen 1981 und 2010. Im östlichen/zentralen Pazifik gab es 26 Stürme, die meisten seit 1992. Im Vergleich hatte der Atlantik weniger Stürme als in den meisten Jahren der letzten zwei Jahrzehnte.

- Die Arktis würde zunehmend wärmer. Die Menge an Seeeis blieb gering. Die 2015 dort gemessenen Oberflächentemperaturen waren 1, 2 °C über dem Schnitt von 1981 - 2010. Damit teilt es sich den ersten Platz des wärmsten Jahres mit 2007 und 2011. Die höchste Konzentration an arktischem Meereis wurde im Februar 2015 gemessen, war aber das Geringste der 37 jährigen Satellitenaufzeichnungen. Die geringste gemessene Meereiskonzentration im September war die viertkleinste seit Beginn der Messungen.

Über den State of the Climate 2015 -Report:
Dieser jährliche "check up"für den Planeten wird vom "National Center for Environmental Information"geleitet und besteht aus Beiträgen von mehr als 450 Wissenschaftlern aus 62 Ländern. Es wurde als spezielle Ergänzung zum "Bulletin of the American Meteorologial Society"veröffentlicht.

English (original): International report confirms Earth is hot and getting hotter
2015 topped 2014 as warmest year on record with help from El Nino.
An annual State of the Climate report has confirmed that 2015 surpassed 2014 as the warmest year on record since at least the mid-to-late 19th century.

Last year's record heat resulted from a combination of long-term global warming and one of the strongest El Nino experienced since at least 1950, according to the more than 450 scientists that contributed to the report. They found that most indicators of climate change continued to reflect trends consistent with a global warming.

Notable findings from the report include:

- Greenhouse gases were the highest on record. Major greenhouse gas concentrations, including carbon dioxide (CO_2), methane and nitrous oxide, rose to new record high values during 2015. The 2015 average global CO_2 concentration was 399.4 parts per million (ppm), an increase of 2.2 ppm compared with 2014.

- Global surface temperature was the highest on record. Aided by the strong El Nino, the 2015 annual global surface temperature was 0.76 - 0.83 °F (0.42 °C - 0.46 °C) above the 1981 - 2010 average, surpassing the previous record set in 2014.

- Sea surface temperature was the highest on record. The globally averaged sea surface temperature was 0.59 - 0.70 °F (0.33 °C - 0.39 °C) above average, breaking the previous mark set in 2014.

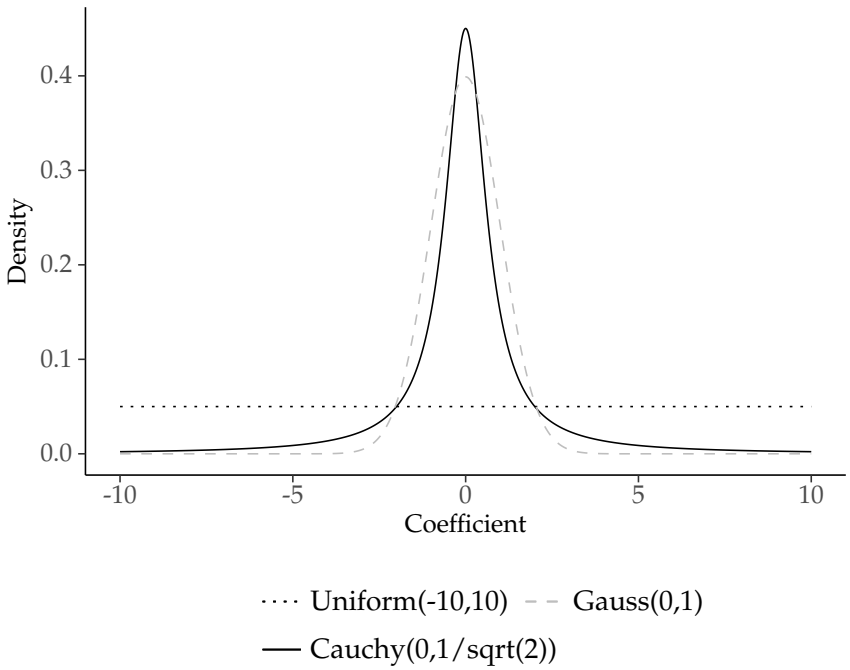
- Global upper ocean heat content highest on record. Upper ocean heat content exceeded the record set in 2014, reflecting the continuing accumulation of heat in the ocean's top layers.

- Global sea level rose to a new record high in 2015. It measured about 2.75 inches (70 mm) higher than that observed in 1993, when satellite record-keeping for global sea level rise began.
- Tropical cyclones were well above average, overall. There were 101 tropical cyclones total across all ocean basins in 2015, well above the 1981 - 2010 average of 82 storms. The eastern/central Pacific had 26 named storms, the most since 1992. The North Atlantic, in contrast, had fewer storms than most years during the last two decades.
- The Arctic continued to warm; sea ice extent remained low. The Arctic land surface temperature in 2015 was 2.2 °F (1.2 °C) above the 1981-2010 average, tying 2007 and 2011 as the highest on record. The maximum Arctic sea ice extent reached in February 2015 was the smallest in the 37-year satellite record, while the minimum sea ice extent that September was the fourth lowest on record.

About the State of the Climate 2015 Report
This yearly "check-up" for the planet, led by NOAA's National Centers for Environmental Information, is based on contributions from more than 450 scientists from 62 countries around the world. It published as a special supplement to the Bulletin of the American Meteorological Society.

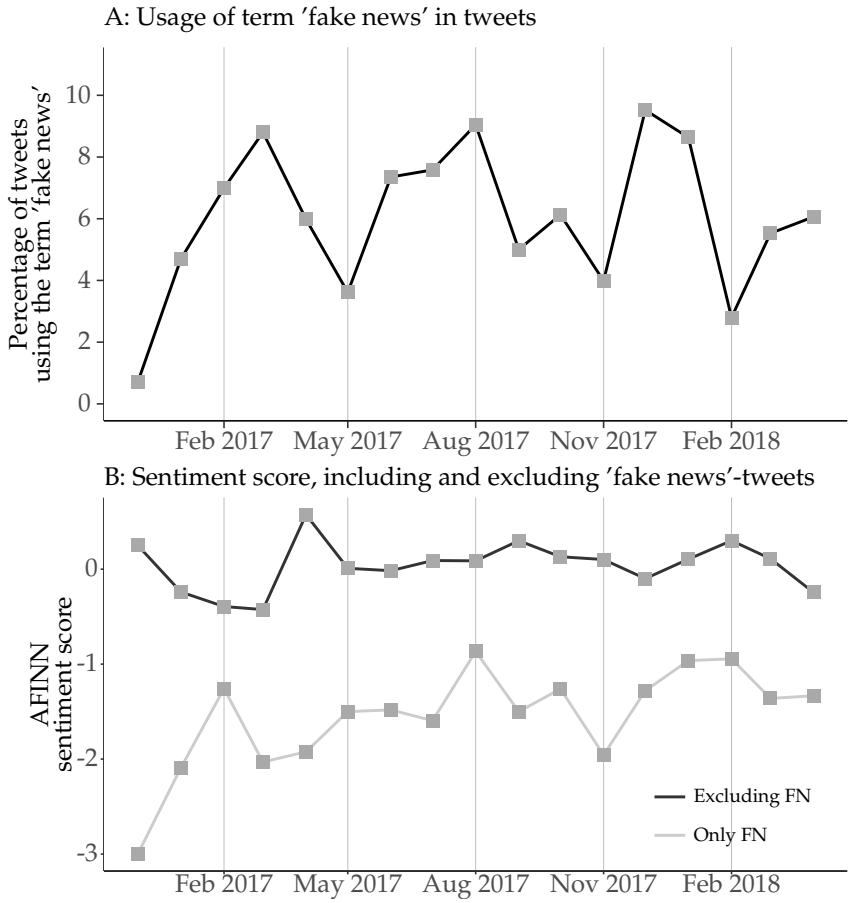
C.2 Statistical analyses

FIGURE C.1: Two prior distributions, and a Gauss distribution for reference



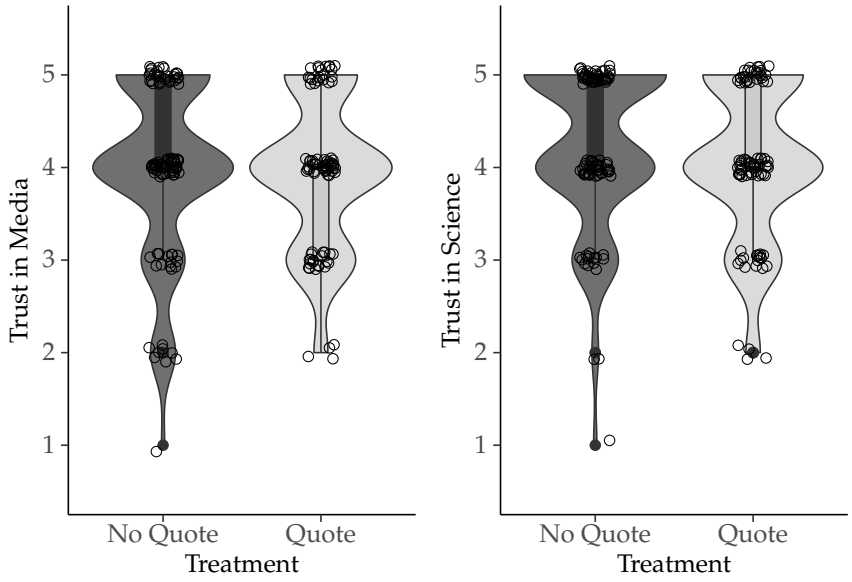
Notes: Shows two prior distributions. The prior with the Cauchy distribution with scale 2 is uninformative, while the Cauchy distribution with scale $\frac{1}{\sqrt{2}}$ incorporates a higher prior belief in the absence of a treatment effect. For reference, a Gaussian distribution with SD of 1 is shown as well.

FIGURE C.2: Trends related to usage of the term 'fake news' by Twitter account of Donald Trump



Notes: Panel A shows the percentage of tweets including the term 'fake news' from his first occurance on. Panel B shows the AFINN sentiment score for all tweets, resp. for only the tweets mentioning 'fake news'. Data is from the Twitter Account @realDonaldTrump. Source: http://www.trumptwitterarchive.com/data/realdonaldtrump/.

FIGURE C.3: Violin plots of trust in media and science



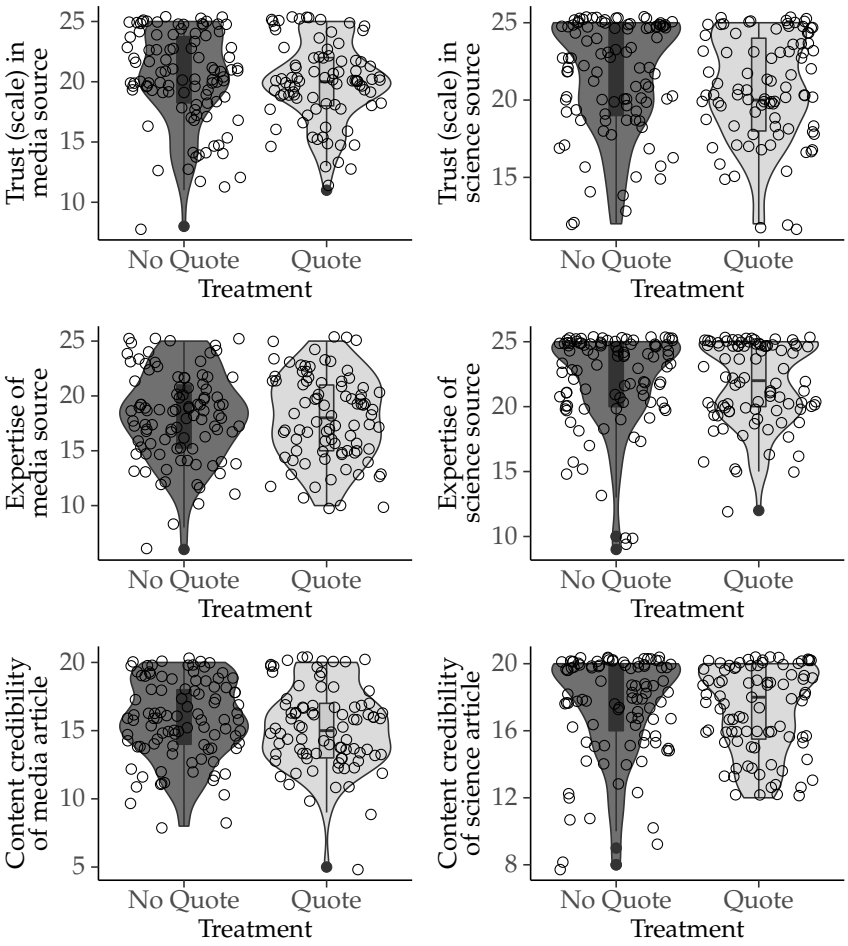
Notes: Shows ordinal level trust in media (left) and science (right) conditional on treatment condition.

TABLE C.1: Ordinal logistic regression of media and science trustworthiness on treatment and covariates

	Media trustworthiness			Science trustworthiness		
	B	CI_{95}	p	B	CI_{95}	p
Treatment						
Quote	-0.201	-0.76 0.35	0.478	-0.405	-0.96 0.15	0.152
Covariates						
Age	-0.039	-0.09 0.02	0.162	-0.029	-0.09 0.03	0.364
Male	-0.224	-0.80 0.35	0.443	-0.318	-0.91 0.28	0.294
Economics	-0.440	-1.16 0.28	0.231	0.100	-0.65 0.85	0.793
History	-0.801	-2.02 0.42	0.199	-0.784	-2.18 0.61	0.269
Law	-0.663	-1.91 0.58	0.296	-1.239	-2.53 0.05	0.060
Management	-0.373	-1.31 0.56	0.435	0.195	-0.72 1.11	0.674
Social Science	0.031	-0.84 0.90	0.944	-0.390	-1.24 0.46	0.366
Philosophy	13.522	-1818.8 1849.6	0.987	16.029	-1769.1 1798.8	0.987
Contribution	0.016	-0.08 0.11	0.730	0.019	-0.07 0.11	0.686
1|2	-6.63	-9.17 -4.10		-6.433	-9.10 -3.77	
2|3	-4.01	-5.71 -2.30		-4.44	-6.39 -2.48	
3|4	-2.46	-4.06 -0.86		-2.51	-4.32 -0.70	
4|5	-0.33	-1.88 1.23		-0.57	-2.33 1.19	
Observations		184			183	
Pseudo R^2		0.0287			0.035	
LR-chi^2(10)		13.00			15.07	
AIC		467.5			441.106	

Notes: Shows point estimates (B), 95 %-confidence intervals (CI_{95}), and p-values (p) from the respective ordinal logistic regression model.

FIGURE C.4: Violin plots of perceptions of trust, expertise, and content credibility in media and science



Notes: Shows ordinal level of secondary outcome scales for media (left) and science (right) conditional on treatment condition.

TABLE C.2: OLS-regression of media trustworthiness, expertise, and content credibility on treatment and covariates

	Media trustworthiness scale			Media expertise scale			Media content credibility		
	B	CI_{95}	p	B	CI_{95}	p	B	CI_{95}	p
Intercept	21.48	18.94 – 24.02	<.001	19.54	15.86 – 23.23	<.001	15.78	13.12 – 18.44	<.001
Treatment									
Quote	-0.33	-1.43 – 0.77	0.553	-0.19	-1.43 – 1.04	0.76	-0.52	-1.41 – 0.36	0.245
Covariates									
Age	-0.03	-0.12 0.06	0.528	-0.05	-0.19 0.08	0.451	0	-0.10 0.09	0.936
Male	-0.39	-1.50 0.73	0.495	-0.6	-1.84 0.64	0.341	-0.57	-1.50 0.35	0.225
Economics	-0.81	-2.15 0.53	0.235	0.29	-1.38 1.95	0.736	-0.14	-1.36 1.08	0.82
History	-2.6	-5.42 0.21	0.07	-1.97	-4.14 0.21	0.076	-1.62	-3.65 0.40	0.116
Law	-1.16	-4.39 2.06	0.477	-0.9	-4.31 2.52	0.605	-0.66	-2.31 1.00	0.435
Management	-0.73	-2.59 1.14	0.443	0.4	-1.46 2.26	0.672	-0.14	-1.84 1.56	0.868
Philosophy	4.51	3.19 – 5.83	<.001	3.63	0.84 6.42	0.011	1.98	-5.71 9.66	0.613
Social Science	0.05	-1.58 1.68	0.955	1.04	-0.78 2.85	0.261	0.71	-0.73 2.15	0.332
Contribution	0.05	-0.12 0.21	0.589	0	-0.20 0.19	0.971	0.12	-0.03 0.27	0.109
Observations		184			184			184	
R^2 / adj. R^2		.065/ .011			.051/ -.004			.070/ .017	
F-statistics		1.212			0.931			1.308	
AIC		1003.039			1040.645			931.385	

Notes: Shows point estimates (B), confidence intervals (CI_{95}), and p-values (p) from the respective OLS regression model. Robust confidence intervals.

TABLE C.3: OLS-regression of science trustworthiness, expertise, and content credibility on treatment and covariates

	Science trustworthiness scale			Science expertise scale			Science content credibility		
	B	CI_{95}	p	B	CI_{95}	p	B	CI_{95}	p
Intercept	22.51	18.40 – 26.62	<.001	22.51	19.31 – 25.72	<.001	17.01	14.40 – 19.62	<.001
Treatment									
Quote	-0.71	-1.76 – 0.34	0.182	-0.27	-1.33 – 0.79	0.617	-0.22	-1.08 – 0.64	0.612
Covariates									
Age	-0.03	-0.18 0.13	0.721	0	-0.12 0.11	0.937	0.01	-0.08 0.10	0.821
Male	-0.72	-1.88 0.44	0.22	-0.44	-1.57 0.70	0.448	-0.39	-1.33 0.56	0.418
Economics	0.35	-1.01 1.72	0.61	-0.56	-2.02 0.91	0.455	0.63	-0.53 1.80	0.286
History	-1.59	-5.18 2.00	0.384	-0.54	-3.06 1.98	0.671	-0.69	-3.64 2.26	0.643
Law	-2.06	-5.41 1.29	0.227	-1.87	-5.21 1.47	0.271	-0.81	-3.31 1.69	0.523
Management	-0.09	-1.79 1.61	0.913	0.45	-1.15 2.04	0.58	0.61	-0.90 2.12	0.427
Philosophy	3.15	0.13 6.18	0.041	2.76	1.51 4.01	<.001	2.77	1.76 3.78	<.001
Social Science	-0.07	-1.58 1.43	0.922	0.52	-0.90 1.95	0.47	0.42	-0.82 1.67	0.505
Contribution	0.04	-0.14 0.22	0.666	0	-0.19 0.18	0.967	0.02	-0.13 0.16	0.824
Observations		183			183			183	
R^2 / adj. R^2		.069/ .015			.043/ -.012			.036/ -.020	
F-statistics		1.269			0.782			0.648	
AIC		985.736			981.385			915.423	

Notes: Shows point estimates (B), confidence intervals (CI_{95}), and p-values (p) from the respective OLS regression model. Robust confidence intervals.

C.3 Questionnaire

- To what extent do you agree with the assertion of the quote? (on a 5-point Likert scale)
  - Not agree at all - Fully agree
- The **source ZEIT ONLINE** appears to me as (on a 5-point Likert differential scale)
  - Undependable - Dependable
  - Dishonest - Honest
  - Unreliable - Reliable
  - Insincere - Sincere
  - Untrustworthy - Trustworthy
  - Not an expert - Expert
  - Inexperienced - Experienced
  - Unknowledgeable - Knowledgeable
  - Unqualified - Qualified
  - Unskilled - Skilled
- The **content** appears to me as (on a 5-point Likert differential scale)
  - Inaccurate - Accurate
  - Unbelievable - Believable
  - Opinionated - Factual
  - Untrustworthy - Trustworthy

CONFIDENTIAL - FOR PEER-REVIEW ONLY

Spillover of distrust - Hamburg, 2017, HB (#5219)

Created: 08/22/2017 06:13 AM (PT)
Shared: 08/22/2017 06:20 AM (PT)

1) What's the main question being asked or hypothesis being tested in this study?

H1: Increasing the salience of doubt in the independence of media reduces perceived trustworthiness of the source of a media article.

H2: Increasing the salience of doubt in the independence of media reduces perceived trustworthiness of the source of a scientific article.

2) Describe the key dependent variable(s) specifying how they will be measured.

Trustworthiness item (5-point semantic differential item) from a semantic differential scale to measure celebrity endorsers' perceived expertise, trustworthiness, and attractiveness (Ohanian, 1990, Journal of Advertising).

3) How many and which conditions will participants be assigned to?

4 conditions in a mixed design laboratory experiment. Participants will be randomly assigned to one of two conditions: subjects will either read a statement expressing doubt in the independence of media, or not read a quote (2 between subjects conditions). After that, all subjects will read two articles that differ with respect to the source, i.e. scientific source, or media source (2 within subjects conditions). The order of articles in the within subject condition is randomized.

4) Specify exactly which analyses you will conduct to examine the main question/hypothesis.

To take into account the ordinal nature of the 5-point semantic differential response items, I use non parametric Chi-squared tests, and Mann Whitney tests to test the null hypotheses that perceived source trustworthiness in the quote condition is equal to or higher than in the no quote condition, both for the media article (H1) and the scientific article (H2). Based on the formulated hypotheses, I conduct one-sided tests.
Additionally, I will estimate the respective ordered logistic regression models with treatment assignment as the respective independent variable, plus control variables (amount contributed to climate protection in the preceding experiment), and socio-demographic variables (field of study, gender, age).

5) Any secondary analyses?

I will also test for treatment effects on two sub-scales of an index measuring perceived source credibility, i.e. perceived source trustworthiness (5 items), and perceived source expertise (5 items) (Ohanian, 1990, Journal of Advertising).
I will also test for effects on perceived content credibility (4 items) (Gaziano and McGrath, 1986, Journalism Quarterly)

6) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.

75 observations per group (no quote/ quote) allow to detect a medium standardized effect size d = 0.42 with 80% power, assuming an alpha level of 0.05 and an independent samples Wilcoxon Mann-Whitney U test. The tested hypotheses are directional:
H0: perceived trustworthiness of the source in the quote-condition is equal to or higher than in the noQuote-condition;
HA: perceived source trustworthiness in the quote-condition is lower than the noQuote-condition.
I will collect observations until having at least 150 observations in total, with equal allocation ratio.

7) Anything else you would like to pre-register? (e.g., data exclusions, variables collected for exploratory purposes, unusual analyses planned?)

In case of non-significant results I will compute Bayes Factors in order to quantify the relative evidence in favor of the null hypothesis and the alternative hypothesis. I will also test for order effects, i.e. whether media, resp. science articles are rated differently with respect to perceptions of trustworthiness depending on their order of being read.
Subjects that take less than the mean time minus two standard deviations to read the respective articles are excluded from the analysis.

8) Have any data been collected for this study already?

No, no data have been collected for this study yet

Bibliography

Agresti, Alan (2007). *An introduction to categorical data analysis*. Hoboken, New Jersey: John Wiley & Sons, Inc.

Alemanno, Alberto and Anne-Lise Sibony (2015). *Nudge and the law: A European perspective*. Oxford and Portland, OR: Bloomsbury Publishing.

Allcott, Hunt and Matthew Gentzkow (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives*, 31 (2): 211–236. DOI: 10.1257/jep.31.2.211.

Allcott, Hunt and Sendhil Mullainathan (2010). Energy Behavior and Energy Policy. *Science*, 327 (5970): 1204–1205. DOI: 10.1126/science.1180775.

Allison, Paul D. (1990). Change Scores as Dependent Variables in Regression Analysis. *Sociological Methodology*, 20: 93–114. DOI: 10.2307/271083.

Altmann, Steffen, Armin Falk, Paul Heidhues, and Rajshri Jayaraman (2014). Defaults and donations: Evidence from a field experiment. Berlin. URL: https://ssrn.com/abstract=2548980.

Altmann, Steffen, Armin Falk, and Andreas Grunewald (2015). Incentives and information as driving forces of default effects. URL: https://ssrn.com/abstract=2329060 (visited on May 15, 2018).

Anderson, Kevin and Glen Peters (2016). The trouble with negative emissions. *Science*, 354 (6309): 182–183. DOI: 10.1126/science.aah4567.

Andreoni, James (1988). Privately provided public goods in a large economy: The limits of altruism. *Journal of Public Economics*, 35 (1): 57–73. DOI: 10.1016/0047-2727(88)90061-8.

Arad, Ayala and Ariel Rubinstein (2017). The People's Perspective on Libertarian-Paternalistic Policies. URL: http://www.tau.ac.il/~aradayal/LP.pdf (visited on May 21, 2018).

Araña, Jorge E. and Carmelo J. León (2013). Can Defaults Save the Climate? Evidence from a Field Experiment on Carbon Offsetting Programs. *Environmental & Resource Economics*, 54 (4): 613–626. DOI: 10.1007/s10640-012-9615-x.

Bakshy, Eytan, Solomon Messing, and Lada A. Adamic (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348 (6239): 1130–1132. DOI: 10.1126/science.aaa1160.

Bang, H. Min, Suzanne B. Shu, and Elke U. Weber (2018). The role of perceived effectiveness on the acceptability of choice architecture. *Behavioural Public Policy*, 11: 1–21. DOI: 10.1017/bpp.2018.1.

Bartling, Björn, Ernst Fehr, and Holger Herz (2014). The Intrinsic Value of Decision Rights. *Econometrica*, 82 (6): 2005–2039. DOI: 10.3982/ECTA11573.

Beard, Emma, Zoltan Dienes, Colin Muirhead, and Robert West (2016). Using Bayes factors for testing hypotheses about intervention effectiveness in

addictions research. *Addiction*, 111 (12): 2230–2247. DOI: 10.1111/add.13501.

Bénabou, Roland and Jean Tirole (2003). Intrinsic and Extrinsic Motivation. *The Review of Economic Studies*, 70 (3): 489–520. DOI: 10.1111/1467-937X.00253.

Bénabou, Roland and Jean Tirole (2006). Incentives and Prosocial Behavior. *The American Economic Review*, 96 (5): 1652–1678. DOI: 10.1257/aer.96.5.1652.

Benartzi, Shlomo, John Beshears, Katherine L. Milkman, Cass R. Sunstein, Richard H. Thaler, Maya Shankar, Will Tucker-Ray, William J. Congdon, and Steven Galing (2017). Should Governments Invest More in Nudging? *Psychological Science*, 28 (8): 1–15. DOI: 10.1177/0956797617702501.

Benjamin, Daniel J. et al. (2017). Redefine statistical significance. *Nature Human Behaviour*, 2: 6–10. DOI: 10.1038/s41562-017-0189-z.

Bock, Olaf, Andreas Nicklisch, and Ingmar Baetge (2014). Hamburg registration and organization online tool. *European Economic Review*, 71: 117–120. DOI: 10.1016/j.euroecorev.2014.07.003.

Bolton, Gary E. and Axel Ockenfels (2012). Behavioral economic engineering. *Journal of Economic Psychology*, 33 (3): 665–676. DOI: 10.1016/j.joep.2011.09.003.

Bordalo, P., N. Gennaioli, and A. Shleifer (2012). Salience Theory of Choice Under Risk. *The Quarterly Journal of Economics*, 127 (3): 1243–1285. DOI: 10.1093/qje/qjs018.

Borghans, Lex and Bart H.H. Golsteyn (2015). Susceptibility to default training options across the population. *Journal of Economic Behavior & Organization*, 117: 369–379. DOI: 10.1016/j.jebo.2015.06.013.

Boussalis, Constantine and Travis G. Coan (2016). Text-mining the signals of climate change doubt. *Global Environmental Change*, 36: 89–100. DOI: 10.1016/j.gloenvcha.2015.12.001.

Bovens, Luc (2009). The Ethics of Nudge. *Preference change*. Ed. by Till Grüne-Yanoff and Sven Ove Hansson. Dordrecht et al.: Springer, pp. 207–219. DOI: 10.1007/978-90-481-2593-7_10.

Bowles, Samuel (2008). Policies Designed for Self-Interested Citizens May Undermine "The Moral Sentiments": Evidence from Economic Experiments. *Science*, 320 (5883): 1605–1609. DOI: 10.1126/science.1152110.

Bowles, Samuel and Sandra Polanía-Reyes (2012). Economic Incentives and Social Preferences: Substitutes or Complements? *Journal of Economic Literature*, 50 (2): 368–425. DOI: 10.1257/jel.50.2.368.

Brehm, Jack W. (1966). *A theory of psychological reactance*. Oxford, England: Academic Press.

Brehm, Sharon S. and Jack W. Brehm (2013). *Psychological reactance: A theory of freedom and control*. New York: Academic Press.

Brown, Christina L. and Aradhna Krishna (2004). The skeptical shopper: A metacognitive account for the effects of default options on choice. *Journal of Consumer Research*, 31 (3): 529–539. DOI: 10.1086/425087.

Bruns, Hendrik, Elena Kantorowicz-Reznichenko, Katharina Klement, Marijane Luistro Jonsson, and Bilel Rahali (2018). Can Nudges Be Transparent

and Yet Effective? *Journal of Economic Psychology*, 65: 41–59. DOI: 10.1016/j.joep.2018.02.002.

Bühren, Christoph and Maria Daskalakis (2015). Do not incentivize eco-friendly behavior: Go for a competition to go green! URL: https://www.econstor.eu/handle/10419/125536 (visited on May 11, 2018).

Bürkner, Paul-Christian (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, 80: 1–28. DOI: 10.18637/jss.v080.i01.

Button, Katherine S., John P. A. Ioannidis, Claire Mokrysz, Brian A. Nosek, Jonathan Flint, Emma S. J. Robinson, and Marcus R. Munafò (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14 (5): 365–376. DOI: 10.1038/nrn3475.

Campbell, Troy H. and Aaron C. Kay (2014). Solution aversion: On the relation between ideology and motivated disbelief. *Journal of Personality and Social Psychology*, 107 (5): 809–824. DOI: 10.1037/a0037963.

Cappelletti, Dominique, Luigi Mittone, and Matteo Ploner (2014). Are default contributions sticky? An experimental analysis of defaults in public goods provision. *Journal of Economic Behavior & Organization*, 108: 331–342. DOI: 10.1016/j.jebo.2014.01.002.

Cardenas, Juan Camilo, John Stranlund, and Cleve Willis (2000). Local environmental control and institutional crowding-out. *World Development*, 28 (10): 1719–1733. DOI: 10.1016/S0305-750X(00)00055-3.

Chaiken, Shelly (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, 39 (5): 752–766. DOI: 10.1037/0022-3514.39.5.752.

Charles, Susan Turk, Chandra A. Reynolds, and Margaret Gatz (2001). Age-related differences and change in positive and negative affect over 23 years. *Journal of Personality and Social Psychology*, 80 (1): 136–151. DOI: 10.1037/0022-3514.80.1.136.

Chartrand, Tanya L., Amy N. Dalton, and Gavan J. Fitzsimons (2007). Nonconscious relationship reactance: When significant others prime opposing goals. *Journal of Experimental Social Psychology*, 43 (5): 719–726. DOI: 10.1016/j.jesp.2006.08.003.

Cho, Hyun-Chul and Shuzo Abe (2013). Is two-tailed testing for directional research hypotheses tests legitimate? *Journal of Business Research*, 66 (9): 1261–1266. DOI: 10.1016/j.jbusres.2012.02.023.

Cook, John, Dana Nuccitelli, Sarah A. Green, Mark Richardson, Bärbel Winkler, Rob Painting, Robert Way, Peter Jacobs, and Andrew Skuce (2013). Quantifying the consensus on anthropogenic global warming in the scientific literature. *Environmental Research Letters*, 8 (2): 24024. DOI: 10.1088/1748-9326/8/2/024024.

Cook, John, Naomi Oreskes, Peter T. Doran, William R. L. Anderegg, Bart Verheggen, Ed W. Maibach, J. Stuart Carlton, Stephan Lewandowsky, Andrew G. Skuce, and Sarah A. Green (2016). Consensus on consensus: a

synthesis of consensus estimates on human-caused global warming. *Environmental Research Letters*, 11 (4): 48002. DOI: 10.1088/1748-9326/11/4/048002.

Costa, Dora L. and Matthew E. Kahn (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.

Croson, R. and M. Marks (2001). The effect of recommended contributions in the voluntary provision of public goods. *Economic Inquiry*, 39 (2): 238–249. DOI: 10.1111/j.1465-7295.2001.tb00063.x.

Croson, Rachel and Uri Gneezy (2009). Gender differences in preferences. *Journal of Economic Literature*, 47 (2): 448–474. DOI: 10.1257/jel.47.2.448.

Croson, Rachel and Nicolas Treich (2014). Behavioral Environmental Economics: Promises and Challenges. *Environmental & Resource Economics*, 58 (3): 335–351. DOI: 10.1007/s10640-014-9783-y.

Daily Sabah (2017). Presidency tops poll as 'most trusted institution' in Turkey. Istanbul. URL: https://www.dailysabah.com/politics/2017/01/19/presidency-tops-poll-as-most-trusted-institution-in-turkey (visited on April 18, 2018).

Davenport, Coral and Eric Lipton (2016). Trump Picks Scott Pruitt, Climate Change Denialist, to Lead E.P.A. URL: https://www.nytimes.com/2016/12/07/us/politics/scott-pruitt-epa-trump.html (visited on April 18, 2018).

Dhingra, Nikhil, Zach Gorn, Andrew Kener, and Jason Dana (2012). The default pull: An experimental demonstration of subtle default effects on preferences. *Judgment and Decision Making*, 7 (1): 69–76. URL: http://journal.sjdm.org/10/10831a/jdm10831a.html.

Di Lorenzo, Giovanni (2017). They Should Look Up What Dictator Means! URL: http://www.zeit.de/politik/ausland/2017-07/recep-tayyip-erdogan-g20-summit-interview-english (visited on May 21, 2018).

Diederich, Johannes and Timo Goeschl (2014). Willingness to Pay for Voluntary Climate Action and Its Determinants: Field-Experimental Evidence. *Environmental & Resource Economics*, 57 (3): 405–429. DOI: 10.1007/s10640-013-9686-3.

Diederich, Johannes and Timo Goeschl (2017). To mitigate or not to mitigate: The price elasticity of pro-environmental behavior. *Journal of Environmental Economics and Management*, 84: 209–222. DOI: 10.1016/j.jeem.2017.03.004.

Dienes, Zoltan (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, 5: 781. DOI: 10.3389/fpsyg.2014.00781.

Dillard, James Price and Lijiang Shen (2005). On the nature of reactance and its role in persuasive health communication. *Communication Monographs*, 72 (2): 144–168. DOI: 10.1080/03637750500111815.

Dinner, Isaac, Eric J. Johnson, Daniel G. Goldstein, and Kaiya Liu (2011). Partitioning default effects: Why people choose not to choose. *Journal of Experimental Psychology: Applied*, 17 (4): 332–341. DOI: 10.1037/a0024354.

Dolan, Paul, M. Hallsworth, D. Halpern, D. King, Robert D. Metcalfe, and I. Vlaev (2012). Influencing behaviour: The mindspace way. *Journal of Economic Psychology*, 33 (1): 264–277. DOI: 10.1016/j.joep.2011.10.009.

Dunlap, Riley E. (2013). Climate Change Skepticism and Denial. *American Behavioral Scientist*, 57 (6): 691–698. DOI: 10.1177/0002764213477097.

Engel, Christoph (2011). Dictator games: a meta study. *Experimental Economics*, 14 (4): 583–610. DOI: 10.1007/s10683-011-9283-7.

Engels, Anita, Otto Hüther, Mike Schäfer, and Hermann Held (2013). Public climate-change skepticism, energy preferences and political participation. *Global Environmental Change*, 23 (5): 1018–1027. DOI: 10.1016/j.gloenvcha.2013.05.008.

Falk, Armin and Michael Kosfeld (2006). The Hidden Costs of Control. *The American Economic Review*, 96 (5): 1611–1630. DOI: 10.2307/30034987.

Faul, Franz, Edgar Erdfelder, Axel Buchner, and Albert-Georg Lang (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41 (4): 1149–1160. DOI: 10.3758/BRM.41.4.1149.

Fehr, Ernst and Urs Fischbacher (2003). The nature of human altruism. *Nature*, 425 (6960): 785–791. DOI: 10.1038/nature02043.

Fehr, Ernst and Simon Gächter (2000). Fairness and Retaliation: The Economics of Reciprocity. *The Journal of Economic Perspectives*, 14 (3): 159–182. DOI: 10.1257/jep.14.3.159.

Fehr, Ernst and Klaus M. Schmidt (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114 (3): 817–868. DOI: 10.1162/003355399556151.

Fehr, Ernst, Holger Herz, and Tom Wilkening (2013). The Lure of Authority: Motivation and Incentive Effects of Power. *The American Economic Review*, 103 (4): 1325–1359. DOI: 10.1257/aer.103.4.1325.

Feldman, Lauren, Teresa A. Myers, Jay D. Hmielowski, and Anthony Leiserowitz (2014). The Mutual Reinforcement of Media Selectivity and Effects: Testing the Reinforcing Spirals Framework in the Context of Global Warming. *Journal of Communication*, 64 (4): 590–611. DOI: 10.1111/jcom.12108.

Felsen, Gidon, Noah Castelo, and Peter B. Reiner (2013). Decisional enhancement and autonomy: Public attitudes towards overt and covert nudges. *Judgment and Decision Making*, 8 (3): 202–213. URL: https://search.proquest.com/docview/1364598265?accountid=11262.

Fischbacher, Urs (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10 (2): 171–178. DOI: 10.1007/s10683-006-9159-4.

Fitzsimons, Gavan J. and Donald R. Lehmann (2004). Reactance to recommendations: When unsolicited advice yields contrary responses. *Marketing Science*, 23 (1): 82–94. DOI: 10.1287/mksc.1030.0033.

Frey, Bruno S. (1992). Pricing and regulating affect environmental ethics. *Environmental & Resource Economics*, 2 (4): 399–414. DOI: 10.1007/BF00304969.

Gaziano, Cecilie and Kristin McGrath (1986). Measuring the concept of credibility. *Journalism Quarterly*, 63 (3): 451–462. URL: http://journals.sagepub.com/doi/10.1177/107769908606300301.

Gelman, Andrew (2008). Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine*, 27 (15): 2865–2873. DOI: 10.1002/sim.3107.

Gelman, Andrew, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin (2014). *Bayesian data analysis*. Boca Raton, FL: Chapman and Hall/CRC press.

Gerber, Alan S. and Donald P. Green (2012). *Field experiments: Design, analysis, and interpretation*. United States: WW Norton.

Gintis, Herbert, Samuel Bowles, Robert Boyd, and Ernst Fehr, eds. (2005). *Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life*. Cambridge: MIT Press.

Goeschl, Timo and Grischa Perino (2012). Instrument Choice and Motivation: Evidence from a Climate Change Experiment. *Environmental & Resource Economics*, 52 (2): 195–212. DOI: 10.1007/s10640-011-9524-4.

Goswami, Indranil and Oleg Urminsky (2016). When Should the Ask Be a Nudge? The Effect of Default Amounts on Charitable Donations. *Journal of Marketing Research*, 53 (5): 829–846. DOI: 10.1509/jmr.15.0001.

Gowdy, John M. (2008). Behavioral economics and climate change policy. *Journal of Economic Behavior & Organization*, 68 (3–4): 632–644. DOI: 10.1016/j.jebo.2008.06.011.

Grüne-Yanoff, Till and Ralph Hertwig (2015). Nudge Versus Boost: How Coherent are Policy and Theory? *Minds and Machines*:DOI: 10.1007/s11023-015-9367-9.

Haggag, Kareem and Giovanni Paci (2014). Default Tips. *American Economic Journal: Applied Economics*, 6 (3): 1–19. DOI: 10.1257/app.6.3.1.

Hagman, William, David Andersson, Daniel Västfjäll, and Gustav Tinghög (2015). Public Views on Policies Involving Nudges. *Review of Philosophy and Psychology*, 6 (3): 439–453. DOI: 10.1007/s13164-015-0263-2.

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. DOI: 10.1017/S1867299X00005468.

Hansen, Pelle Guldborg and Andreas Maaloe Jespersen (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.

Harrison, Glenn W. and John A. List (2004). Field Experiments. *Journal of Economic Literature*, 42 (4): 1009–1055. DOI: 10.1257/0022051043004577.

Hausman, Daniel M. and Brynn Welch (2010). Debate: To Nudge or Not to Nudge. *Journal of Political Philosophy*, 18 (1): 123–136. DOI: 10.1111/j.1467-9760.2009.00351.x.

Hedlin, Simon and Cass R. Sunstein (2016). Does active choosing promote green energy use? Experimental evidence. *Ecology Law Quarterly*, 43 (1): 107–142. DOI: 10.2139/ssrn.2624359.

Heino, Matti T. J., Matti Vuorre, and Nelli Hankonen (2018). Bayesian evaluation of behavior change interventions: A brief introduction and a practical example. *Advanced Methods in Health Psychology and Behavioral Medicine*, 6 (1): 49–78. DOI: 10.1080/21642850.2018.1428102.

Hmielowski, Jay D., Lauren Feldman, Teresa A. Myers, Anthony Leiserowitz, and Edward Maibach (2014). An attack on science? Media use, trust in scientists, and perceptions of global warming. *Public Understanding of Science*, 23 (7): 866–883. DOI: 10.1177/0963662513480091.

Hoenig, John M. and Dennis M. Heisey (2001). The Abuse of Power. *The American Statistician*, 55 (1): 19–24. DOI: 10.1198/000313001300339897.

Hong, S.-M. and S. Faedda (1996). Refinement of the Hong psychological reactance scale. *Educational and Psychological Measurement*, 56 (1): 173–182. DOI: 10.1177/0013164496056001014.

House of Lords Report (2011). Behaviour Change. Ed. by House of Lords. London. URL: https://publications.parliament.uk/pa/ld201012/ldselect/ldsctech/179/17902.htm (visited on April 18, 2018).

Ioannidis, John P. A., Tom D. Stanley, and Hristos Doucouliagos (2017). The power of bias in economics research. *The Economic Journal*, 127 (605): F236–F265. DOI: 10.1111/ecoj.12461.

IPCC (2014). Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Ed. by Core Writing Team, R.K. Pachauri, L.A. Meyer. IPCC, Geneva, Switzerland. URL: https://www.ipcc.ch/report/ar5/syr/ (visited on April 18, 2018).

Jachimowicz, Jon M., Shannon Duncan, Elke U. Weber, and Eric J. Johnson (2017). When and Why Defaults Influence Decisions: A Meta-Analysis of Default Effects. DOI: 10.2139/ssrn.2727301.

John, Leslie K., George Loewenstein, and Drazen Prelec (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 23 (5): 524–532. URL: http://journals.sagepub.com/doi/10.1177/0956797611430953.

Johnson, Eric J. and Daniel G. Goldstein (2003). Do defaults save lives? *Science*, 302: 1338–1339. DOI: 10.1126/science.1091721.

Kahan, Dan M. (2010). Fixing the communications failure. *Nature*, 463 (7279): 296–297. DOI: 10.1038/463296a.

Kahan, Dan M. (2017). Misconceptions, Misinformation, and the Logic of Identity-Protective Cognition. DOI: 10.2139/ssrn.2973067.

Kahan, Dan M., Donald Braman, John Gastil, Paul Slovic, and C. K. Mertz (2007). Culture and identity–protective cognition: Explaining the white–male effect in risk perception. *Journal of Empirical Legal Studies*, 4 (3): 465–505. DOI: 10.1111/j.1740-1461.2007.00097.x.

Kahan, Dan M., Ellen Peters, Maggie Wittlin, Paul Slovic, Lisa Larrimore Ouellette, Donald Braman, and Gregory Mandel (2012). The polarizing impact of science literacy and numeracy on perceived climate change risks. *Nature Climate Change*, 2 (10): 732–735. DOI: 10.1038/nclimate1547.

Kahan, Dan M., Ellen Peters, Erica Dawson, and Paul Slovic (2017). Motivated numeracy and enlightened self-government. *Behavioural Public Policy*, 1 (1): 54–86. DOI: 10.1017/bpp.2016.2.

Katz, Elihu, Paul Felix Lazarsfeld, and Elmo Roper (2005). *Personal Influence: The part played by people in the flow of mass communications*. New York: Routledge.

Knobloch-Westerwick, Silvia, Cornelia Mothes, Benjamin K. Johnson, Axel Westerwick, and Wolfgang Donsbach (2015). Political Online Information Searching in Germany and the United States: Confirmation Bias, Source Credibility, and Attitude Impacts. *Journal of Communication*, 65 (3): 489–511. DOI: 10.1111/jcom.12154.

Knobloch-Westerwick, Silvia, Cornelia Mothes, and Nick Polavin (2017). Confirmation Bias, Ingroup Bias, and Negativity Bias in Selective Exposure to Political Information. *Communication Research*, 00 (0): 1–21. DOI: 10.1177/0093650217719596.

Kollmuss, Anja and Julian Agyeman (2002). Mind the gap: Why do people act environmentally and what are the barriers to pro-environmental behavior? *Environmental Education Research*, 8 (3): 239–260. DOI: 10.1080/13504620220145401.

Kraft-Todd, Gordon, Erez Yoeli, Syon Bhanot, and David Rand (2015). Promoting cooperation in the field. *Current Opinion in Behavioral Sciences*, 3: 96–101. DOI: 10.1016/j.cobeha.2015.02.006.

Kroese, Floor M., David R. Marchiori, and Denise T. D. de Ridder (2016). Nudging healthy food choices: a field experiment at the train station. *Journal of Public Health*, 38 (2): e133–e137. DOI: 10.1093/pubmed/fdv096.

Kruschke, John (2015). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*. London et al.: Academic Press.

Kuang, Xi, Roberto A. Weber, and Jason Dana (2007). How effective is advice from interested parties? An experimental test using a pure coordination game. *Journal of Economic Behavior & Organization*, 62 (4): 591–604. DOI: 10.1016/j.jebo.2005.03.010.

Kunda, Ziva (1990). The case for motivated reasoning. *Psychological Bulletin*, 108 (3): 480–498. DOI: 10.1037/0033-2909.108.3.480.

Lakens, D., N. McLatchie, P. M. Isager, A. M. Scheel, and Z. Dienes (2018a). Improving Inferences about Null Effects with Bayes Factors and Equivalence Tests. DOI: 10.17605/OSF.IO/QTZWR. (Visited on May 21, 2018).

Lakens, D., F. Adolfi, C. Albers, F. Anvari, M. Apps, S. Argamon, T. Baguley, R. Becker, S. Benning, and D. Bradford (2018b). Justify your alpha. *Nature Human Behavior*, 2: 168–171. DOI: 10.1038/s41562-018-0311-x.

Lange, Andreas, Claudia Schwirplies, and Andreas Ziegler (2014). On the interrelation between carbon offsetting and other voluntary climate protection activities: Theory and empirical evidence. URL: http://hdl.handle.net/10419/105083 (visited on May 12, 2018).

Lazer, David M. J. et al. (2018). The science of fake news. *Science*, 359 (6380): 1094–1096. DOI: 10.1126/science.aao2998.

Lee, Tien-Tsung (2005). The Liberal Media Myth Revisited: An Examination of Factors Influencing Perceptions of Media Bias. *Journal of Broadcasting & Electronic Media*, 49 (1): 43–64. DOI: 10.1207/s15506878jobem4901_4.

Leviston, Zoe, Ian Walker, and S. Morwinski (2013). Your opinion on climate change might not be as common as you think. *Nature Climate Change*, 3 (4): 334. DOI: 10.1038/nclimate1743.

Levitt, Steven D. and John A. List (2009). Field experiments in economics: The past, the present, and the future. *European Economic Review*, 53 (1): 1–18. DOI: 10.1016/j.euroecorev.2008.12.001.

Loewenstein, George, Cindy Bryce, David Hagmann, and Sachin Rajpal (2015). Warning: You Are About To Be Nudged. *Behavioral Science & Policy*, 1 (1): 35–42. DOI: 10.1353/bsp.2015.0000.

Lourenco, Joana Sousa, Emanuele Ciriolo, Sara Rafael Almeida, and Xavier Troussard (2016). Behavioural insights applied to policy: European Report 2016. DOI: 10.2760/04480. URL: https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/behavioural-insights-applied-policy-european-report-2016 (visited on April 18, 2018).

Madrian, Brigitte C. and Dennis F. Shea (2001). The power of suggestion: Inertia in 401 (k) participation and savings behavior. *The Quarterly Journal of Economics*, 116 (4): 1149–1187. DOI: 10.1162/003355301753265543.

Maniadis, Zacharias, Fabio Tufano, and John A. List (2017). To Replicate or Not To Replicate? Exploring Reproducibility in Economics through the Lens of a Model and a Pilot Study. *The Economic Journal*, 127 (605): F209–F235. DOI: 10.1111/ecoj.12527.

McCright, Aaron M. and Riley E. Dunlap (2011). Cool dudes: The denial of climate change among conservative white males in the United States. *Global Environmental Change*, 21 (4): 1163–1172. DOI: 10.1016/j.gloenvcha.2011.06.003.

McKenzie, Craig R. M., Michael J. Liersch, and Stacey R. Finkelstein (2006). Recommendations Implicit in Policy Defaults. *Psychological Science*, 17 (5): 414–420. DOI: 10.1111/j.1467-9280.2006.01721.x.

McShane, Blakeley B., David Gal, Andrew Gelman, Christian Robert, and Jennifer L. Tackett (2018). Abandon Statistical Significance. URL: https://arxiv.org/abs/1709.07588 (visited on April 18, 2018).

Meier, Stephan (2007). Do subsidies increase charitable giving in the long run? Matching donations in a field experiment. *Journal of the European Economic Association*, 5 (6): 1203–1222. DOI: 10.1162/JEEA.2007.5.6.1203.

Milman, Oliver (2017). US federal department is censoring use of term 'climate change', emails reveal. URL: https://www.theguardian.com/environment/2017/aug/07/usda-climate-change-language-censorship-emails (visited on April 18, 2018).

Morey, Richard D., Jeffrey N. Rouder, and T. Jamil (2015). BayesFactor: Computation of Bayes factors for common designs. URL: https://CRAN.R-project.org/package=BayesFactor (visited on May 21, 2018).

Moser, Susanne C. (2010). Communicating climate change: history, challenges, process and future directions. *WIREs Climate Change*, 1 (1): 31–53. DOI: 10.1002/wcc.11.

National Oceanic and Atmospheric Administration (2016). International report confirms Earth is hot and getting hotter: 2015 topped 2014 as warmest year on record with help from El Nino. URL: http://www.noaa.gov/news/international-report-confirms-earth-is-hot-and-getting-hotter (visited on April 18, 2018).

Nature Human Behaviour (2017). Promoting reproducibility with registered reports. *Nature Human Behaviour*, 1 (1): 0034. DOI: 10.1038/s41562-016-0034.

Newman, Todd P. (2017). Tracking the release of IPCC AR5 on Twitter: Users, comments, and sources following the release of the Working Group I Summary for Policymakers. *Public Understanding of Science*, 26 (7): 815–825. DOI: 10.1177/0963662516628477.

Nickerson, Raymond S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2 (2): 175–220. DOI: 10.1037/1089-2680.2.2.175.

Nielsen, Finn Årup (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. URL: https://arxiv.org/abs/1103.2903 (visited on May 21, 2018).

Nosek, Brian A., Charles R. Ebersole, Alexander C. DeHaven, and David T. Mellor (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences of the United States of America*, 115 (11): 2600–2606. DOI: 10.1073/pnas.1708274114.

Ohanian, Roobina (1990). Construction and validation of a scale to measure celebrity endorsers' perceived expertise, trustworthiness, and attractiveness. *Journal of Advertising*, 19 (3): 39–52. DOI: 10.1080/00913367.1990.10673191.

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, 349 (6251): aac4716. DOI: 10.1126/science.aac4716.

Owens, David, Zachary Grossman, and Ryan Fackler (2014). The control premium: A preference for payoff autonomy. *American Economic Journal: Microeconomics*, 6 (4): 138–161. DOI: 10.1257/mic.6.4.138.

Page, Benjamin I., Robert Y. Shapiro, and Glenn R. Dempsey (1987). What Moves Public Opinion? *American Political Science Review*, 81 (1): 23–43. DOI: 10.2307/1960777.

Perino, Grischa (2015). Climate campaigns, cap and trade, and carbon leakage: Why trying to reduce your carbon footprint can harm the climate. *Journal of the Association of Environmental and Resource Economists*, 2 (3): 469–495. DOI: 10.1086/682572.

Perino, Grischa, Luca A. Panzone, and Timothy Swanson (2014). Motivation crowding in real consumption decisions: Who is messing with my groceries? *Economic Inquiry*, 52 (2): 592–607. DOI: 10.1111/ecin.12024.

Quattrociocchi, Walter, Antonio Scala, and Cass R. Sunstein (2016). Echo Chambers on Facebook. URL: https://ssrn.com/abstract=2795110 (visited on May 21, 2018).

Rebonato, Riccardo (2014). A Critical Assessment of Libertarian Paternalism. *Journal of Consumer Policy*, 37 (3): 357–396. DOI: 10.1007/s10603-014-9265-1.

Reeson, Andrew F. and John G. Tisdell (2008). Institutions, motivations and public goods: An experimental test of motivational crowding. *Journal of Economic Behavior & Organization*, 68 (1): 273–281. DOI: 10.1016/j.jebo.2008.04.002.

Reisch, Lucia A. and Cass R. Sunstein (2016). Do Europeans Like Nudges? *Judgment and Decision Making*, 11 (4): 310–325. URL: http://journal.sjdm.org/16/16202b/jdm16202b.html.

Robinson, David (2017). Trump's Android and iPhone tweets, one year later. URL: http://varianceexplained.org/r/trump-followup/ (visited on April 18, 2018).

Rode, Julian, Erik Gómez-Baggethun, and Torsten Krause (2015). Motivation crowding by economic incentives in conservation policy: A review of the empirical evidence. *Ecological Economics*, 117: 270–282. DOI: 10.1016/j.ecolecon.2014.09.029.

Rogelj, Joeri, Michel den Elzen, Niklas Höhne, Taryn Fransen, Hanna Fekete, Harald Winkler, Roberto Schaeffer, Fu Sha, Keywan Riahi, and Malte Meinshausen (2016). Paris Agreement climate proposals need a boost to keep warming well below 2 °C. *Nature*, 534 (7609): 631–639. DOI: 10.1038/nature18307.

Samuelson, William and Richard Zeckhauser (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty*, 1 (1): 7–59. DOI: 10.1007/BF00055564.

Schäfer, M. S. and I. Schlichting (2014). Media Representations of Climate Change: A Meta-Analysis of the Research Field. *Environmental Communication*, 8 (2): 142–160. DOI: 10.1080/17524032.2014.914050.

Schimmack, Ulrich (2018). The Deductive Fallacy in (some) Bayesian Inductive Inferences. URL: https://replicationindex.wordpress.com/2018/03/03/the-deductive-fallacy-in-some-bayesians-inductive-inferences/ (visited on May 9, 2018).

Schubert, Christian (2017). Green nudges: Do they work? Are they ethical? *Ecological Economics*, 132: 329–342. DOI: 10.1016/j.ecolecon.2016.11.009.

Shear, Michael D. (2017). Trump Will Withdraw U.S. From Paris Climate Agreement. URL: https://www.nytimes.com/2017/06/01/climate/trump-paris-climate-agreement.html (visited on April 18, 2018).

Shogren, J. F. and L. O. Taylor (2008). On Behavioral-Environmental Economics. *Review of Environmental Economics and Policy*, 2 (1): 26–44. DOI: 10.1093/reep/rem027.

Simmons, Joseph P., Leif D. Nelson, and Uri Simonsohn (2011). False-positive psychology: undisclosed flexibility in data collection and analysis allows

presenting anything as significant. *Psychological Science*, 22 (11): 1359–1366. DOI: 10.1177/0956797611417632.

Simon, Herbert A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, 69 (1): 99–118. DOI: 10.2307/1884852.

Simon, Herbert A. (1990). Bounded Rationality. *Utility and Probability*. Ed. by Eatwell J., Milgate M., and Newman P. The New Palgrave. London: Springer, pp. 15–18. DOI: 10.1007/978-1-349-20568-4.

Steffel, Mary, Elanor F. Williams, and Ruth Pogacar (2016). Ethically Deployed Defaults: Transparency and Consumer Protection Through Disclosure and Preference Articulation. *Journal of Marketing Research*, 53 (5): 865–880. DOI: 10.1509/jmr.14.0421.

Stern, Nicholas (2008). The economics of climate change. *The American Economic Review*, 98 (2): 1–37. DOI: 10.1257/aer.98.2.1.

Stroud, Natalie Jomini (2008). Media Use and Political Predispositions: Revisiting the Concept of Selective Exposure. *Political Behavior*, 30 (3): 341–366. DOI: 10.1007/s11109-007-9050-9.

Sugden, Robert (1984). Reciprocity: the supply of public goods through voluntary contributions. *The Economic Journal*, 94 (376): 772–787. DOI: 10.2307/2232294.

Sunstein, Cass R. (2014a). *Simpler: The future of government*. New York: Simon and Schuster.

Sunstein, Cass R. (2014b). *Why nudge? The politics of Libertarian Paternalism*. New Haven, CN: Yale University Press.

Sunstein, Cass R. (2015). Fifty Shades of Manipulation. *Journal of Behavioral Marketing*, 1: 213–244. DOI: 10.2139/ssrn.2565892.

Sunstein, Cass R. (2016). Autonomy by Default. *American Journal of Bioethics*, 16 (11): 1–2. DOI: 10.2139/ssrn.2853750.

Sunstein, Cass R. (2018). Misconceptions about nudges. *Journal of Behavioral Economics for Policy*, 2 (1): 61–67. DOI: 10.2139/ssrn.3033101.

Sunstein, Cass R. and Lucia A. Reisch (2016). Behaviorally Green: Why, Which and When Defaults Can Help. *New Perspectives for Environmental Policies through Behavioral Economics*. Ed. by Frank Beckenbach and Walter Kahlenborn. Heidelberg et al.: Springer, pp. 161–194. DOI: 10.1007/978-3-319-16793-0_7.

Swift, Art (2016). Americans' Trust in Mass Media Sinks to New Low. URL: http://www.gallup.com/poll/195542/americans-trust-mass-media-sinks-new-low.aspx (visited on April 16, 2018).

Tannenbaum, David, Craig Fox, and Todd Rogers (2017). On the misplaced politics of behavioral policy interventions. *Nature Human Behaviour*, 1: 1–7. DOI: 10.1038/s41562-017-0130.

Thaler, Richard H. (2016). Behavioral Economics: Past, Present, and Future. *The American Economic Review*, 106 (7): 1577–1600. DOI: 10.1257/aer.106.7.1577.

Thaler, Richard H. and Cass R. Sunstein (2008). *Nudge: Improving decisions about health, wealth, and happiness*. New Haven, CN: Yale University Press.

Tollefson, Jeff (2010). Climate science: An erosion of trust? *Nature*, 466 (7302): 24–26. DOI: 10.1038/466024a.

Turcotte, Jason, Chance York, Jacob Irving, Rosanne M. Scholl, and Raymond J. Pingree (2015). News Recommendations from Social Media Opinion Leaders: Effects on Media Trust and Information Seeking. *Journal of Computer-Mediated Communication*, 20 (5): 520–535. DOI: 10.1111/jcc4.12127.

UNFCCC (2015). Adoption of the Paris Agreement. URL: https://unfccc.int/resource/docs/2015/cop21/eng/l09r01.pdf (visited on April 18, 2018).

van Vuuren, Detlef P. et al. (2018). Alternative pathways to the 1.5 °C target reduce the need for negative emission technologies. *Nature Climate Change*, 5: 519. DOI: 10.1038/s41558-018-0119-8.

Wagenmakers, Eric-Jan, Tom Lodewyckx, Himanshu Kuriyal, and Raoul Grasman (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage-Dickey method. *Cognitive Psychology*, 60 (3): 158–189. DOI: 10.1016/j.cogpsych.2009.12.001.

Wagenmakers, Eric-Jan, Ruud Wetzels, Denny Borsboom, and Van Der Maas, Han LJ (2011). Why psychologists must change the way they analyze their data: the case of psi: comment on Bem (2011). *Journal of Personality and Social Psychology*, 100 (3): 426–432. DOI: 10.1037/a0022790.

Waldman, Simon A. and Emre Caliskan (2017). The archipelago of press restriction in Turkey. URL: http://eprints.lse.ac.uk/81643/ (visited on April 18, 2018).

Walter, Stefanie, Michael Brüggemann, and Sven Engesser (2018). Echo Chambers of Denial: Explaining user comments on climate change. *Environmental Communication*, 12 (2): 204–217. DOI: 10.1080/17524032.2017.1394893.

Weber, Elke U. and Eric J. Johnson (2012). Psychology and Behavioral Economics Lessons for the Design of a Green Growth Strategy. URL: https://ssrn.com/abstract=2165678 (visited on May 21, 2018).

Weber, Elke U. and Paul C. Stern (2011). Public understanding of climate change in the United States. *American Psychologist*, 66 (4): 315. DOI: 10.1037/a0023253.

Wilson, Timothy D., Christopher E. Houston, Kathryn M. Etling, and Nancy Brekke (1996). A New Look at Anchoring Effects: Basic Anchoring and Its Antecedents. *Journal of Experimental Psychology: General*, 125 (4): 387. URL: http://psycnet.apa.org/buy/1996-06578-004.

World Bank (2015). World Development Report 2015: Mind, Society, and Behavior. Washington, D.C. URL: http://worldbank.org/en/publication/wdr2015 (visited on April 18, 2018).

Wynes, Seth and Kimberly Nicholas (2017). The climate mitigation gap: education and government recommendations miss the most effective individual actions. *Environmental Research Letters*, 12 (7): 74024. DOI: 10.1088/1748-9326/aa7541.

Zeit Online (2016). Dürren, Zyklone, schrumpfende Gletscher. URL: http://www.zeit.de/wissen/umwelt/2016-08/klima-wetter-rekord-el-nino (visited on April 18, 2018).

Zhou, Min (2014). Public environmental skepticism: A cross-national and multilevel analysis. *International Sociology*, 30 (1): 61–85. DOI: 10.1177/0268580914558285.

Ziegler, Andreas (2017). Political orientation, environmental values, and climate change beliefs and attitudes: An empirical cross country analysis. *Energy Economics*, 63: 144–153. DOI: 10.1016/j.eneco.2017.01.022.