Article:

Galizzi, Matteo M. and Whitmarsh, Lorraine (2019) How to measure behavioural spillovers: a methodological review and checklist. Frontiers in Psychology, 10. ISSN 1664-1078

https://doi.org/10.3389/fpsyg.2019.00342

HOW TO MEASURE BEHAVIOURAL SPILLOVERS? A METHODOLOGICAL REVIEW AND CHECKLIST

Matteo M Galizzi* (London School of Economics) and Lorraine Whitmarsh (Cardiff University)
* Corresponding author: Dr. Matteo M Galizzi (m.m.galizzi@lse.ac.uk)

Accepted for publication in *Frontiers in Psychology*.

Abstract
A growing stream of literature at the interface between economics and psychology is currently investigating 'behavioural spillovers' in (and across) different domains, including health, environmental, and pro-social behaviours. A variety of empirical methods have been used to measure behavioural spillovers to date, from qualitative self-reports to statistical/econometric analyses, from online and lab experiments to field experiments. The aim of this paper is to critically review the main experimental and non-experimental methods to measure behavioural spillovers to date, and to discuss their methodological strengths and weaknesses. A consensus mixed-method approach is then discussed which uses between-subjects randomisation and behavioural observations together with qualitative self-reports in a longitudinal design in order to follow up subjects over time. In particular, participants to an experiment are randomly assigned to a treatment group where a behavioural intervention takes place to target behaviour 1, or to a control group where behaviour 1 takes place absent any behavioural intervention. A behavioural spillover is empirically identified as the effect of the behavioural intervention in the treatment group on a subsequent, not targeted, behaviour 2, compared to the corresponding change in behaviour 2 in the control group. Unexpected spillovers and additional insights (e.g., drivers, barriers, mechanisms) are elicited through analysis of qualitative data. In the spirit of the pre-analysis plan, a systematic checklist is finally proposed to guide researchers and policy-makers through the main stages and features of the study design in order to rigorously test and identify behavioural spillovers, and to favour transparency, replicability, and meta-analysis of studies.

Keywords: spillover, mixed-methods, experimental design

Author contributions: MMG initiated and led the paper writing. LW contributed to paper writing.

Acknowledgements: Funding for Whitmarsh was received from the European Research Council, CASPI Starting Grant (336665).

1 Introduction

1.1 What does spillover offer?

Academic and policy interest in 'behavioural spillover' has grown considerably in recent years (e.g., Austin et al., 2011; Truelove et al., 2014; Nilsson et al., 2016). Spillover is where the adoption of one behaviour causes the adoption of additional, related behaviours. As we discuss below, we assume that the initial behaviour change is due to an intervention, although other definitions of behavioural spillovers do not assume this (Nash et al., 2017). From a policy or practitioner perspective, the notion of behavioural spillover is attractive because it appears to hold the promise of changing a suite of behaviours in a cost-effective manner with little regulation which might be politically unpopular. For many pressing social issues, such as climate change or obesity, spillover is thus a promising method of achieving the scale of lifestyle change required to address these, in contrast to the typically small-scale behavioural changes achieved from most individually-focussed interventions (Capstick et al., 2014). From an academic perspective, spillover is intriguing because it sheds new light on the process of lifestyle change: rather than examining behaviour change from the perspective of individual behaviours in isolation, spillover draws attention to the holistic relationships between behaviours within and between contexts, and hence refocus the researchers' perspective on the complex behavioural ecologies that represent lifestyles (Geller, 2001; Schatzki, 2010).

A variety of empirical methods have been used to measure behavioural spillovers to date, from qualitative self-reports to statistical/econometric analyses, from online and lab experiments to field experiments. Detecting spillover has often proved challenging, and there is a need for both conceptual and methodological clarity in order to move the field forward. The aim of this paper is to critically review the main experimental and non-experimental methods to measure behavioural spillovers to date, and to discuss their methodological strengths and weaknesses. A consensus mixed-method approach is then discussed which uses between-subjects randomisation and behavioural observations together with qualitative self-reports in a longitudinal design in order to follow up subjects over time. We conclude by proposing a systematic checklist to guide researchers and policy-makers through the main stages and features of the study design in order to rigorously test and identify behavioural spillovers, and to favour transparency, replicability, and meta-analysis of studies.

1.2 Definition of behavioural spillover

The term 'spillover' has been applied to a wide variety of phenomena, including the spread of knowledge, attitudes, roles/identities, or behaviours from a given domain (e.g., health, environment, care-giving), group, or location, to a different domain, group or location (e.g., Geller, 2001; Poroli & Huang, 2018; Rodriguez-Muñoz et al., 2014; Littleford et al., 2014; Poortinga et al., 2013). The main appeal of such broad definition of behavioural spillover is that it encompasses a rich variety of spillover effects at both a micro and a macro level which are of key interest for policy and practice purposes, such as cross-domains, inter-personal, and cross-regional spillover effects of phenomena and interventions. However, the processes underpinning these diverse effects are highly heterogeneous, ranging from cognition (e.g., learning, problem-solving) and self-regulation, through interpersonal effects (e.g., modelling, contagion) to individual behaviour change, and there is little these processes have in common besides the idea of (often unanticipated) diffusion of some effect.

In what follows, we assume a narrower and more specific definition of behavioural spillover that matches more closely the methodological approach that we have in mind. In particular, behavioural spillover can be defined as the observable and causal effect that a change in one behaviour (behaviour 1) has on a different, subsequent behaviour (behaviour 2). More specifically, to constitute behavioural spillover, the two behaviours must be different (i.e., not related components of a single behaviour), sequential (i.e., behaviour 2 follows behaviour 1), and sharing, at a conscious or unconscious way, an underlying motive (i.e., an overarching goal or a 'deep preference', such as, for example, pro-environmentalism or a healthy life) (Dolan and Galizzi, 2015; Nash, Whitmarsh, Capstick, Hargreaves, Poortinga, Thomas, Sautkina, and Xenias, 2017). This concept of spillover has been examined in relation to different domains (safety, environment, health, finances, etc.) for some decades, although these effects have previously been labelled in diverse ways, including 'response generalisation' (Geller, 2001; Ludwig & Geller, 1997), 'the foot in the door effect' (Freedman & Fraser, 1966; Beaman et al., 1983), and 'moral licensing' (Blanken et al., 2016; Mullen & Monin, 2016). We have conducted a systematic review of the literature (see Appendix A for full details) and found that a total of 106 studies to date have used the above, more specific, definition of behavioural spillovers.[1]

Behavioural spillovers can be categorised as 'promoting', 'permitting', 'purging', or 'precipitating', as illustrated in Figure 1.

Figure 1: Types of behavioural spillovers (adapted from Dolan & Galizzi, 2015: no copyright permissions are required for the reproduction of this figure): examples from health behaviour

	Behaviour 2		
Behaviour 1		*Eat healthily*	*Eat less healthily*
	A run after work	**Promoting**	**Permitting**
		I ran an hour, let's keep up the good work	I ran an hour, I deserve a big slice of cake
	Sofa-sitting after work	**Purging**	**Precipitating**

[1] These 106 studies are: Alpizar et al. (2013a,b); Angelovski et al. (2018); Baca-Motes et al. (2013); Banerjee (2016); Bech-Larsen and Kazbare (2014); Bednar and Page (2018); Bednar et al. (2012); Bednar et al. (2015); Bratt (1999); Capstick et al. (2019); Carpenter and Lawler (2017); Carrico et al. (2017); Cason and Gangadharan (2013); Cason et al. (2012); Chatelain et al. (2018); Claes and Miliute-Plepiene (2018); Cornelissen et al. (2008); Crookes (2017); Dickinson and Oxoby (2011); Dittmer and Blazejewski (2016); Dolan and Galizzi (2014, 2015); Dutschke et al. (2018); Eby (2016); Falk et al. (2013); Fanghella et al. (2019); Fenger (2017); Galbiati et al. (2017); Ghesla et al. (2018); Gholamzadehmir (2016); Godoy et al. (2013); Goswami and Urminsky (2015); Ha and Kwon (2016); Hecht and Boies (2009); Hedrick et al. (2017); Hertwich (2005); Jessoe et al. (2017); Juhl et al. (2017); Juvina et al. (2013); Kaida and Kaida (2015); Karmarkar and Bollinger (2015); Karremans et al. (2005); Kesternich et al. (2017); Klein (2017); Krieg and Samek (2017); Krpan et al. (2019); Lacasse (2015, 2016, 2017); Lanzini and Thøgersen (2014); Lauren et al. (2016, 2017); Lawler (2018); Littleford et al. (2014); Liu et al. (2018); Margetts and Kashima (2016); McCoy and Lyons (2017); Nash et al. (2017, 2018); Nilsson et al. (2016); Nolan (2011); Norden (2013); Panos (2018); Peters et al. (2018); Polizzi di Sorrentino et al. (2016); Poortinga et al. (2013); Santarius and Soland (2018); Savikhin and Sheremeta (2013); Savikhin (2010); Schmitz (2018); Schütte and Gregory-Smith (2015); Seebauer (2018); Sheremeta et al. (2010); Shreedhar and Mourato (2018); Shreedhar (2018); Sintov et al. (2019); Sorrell et al. (2009); Spence et al. (2014); Steinhorst and Matthies (2016); Steinhorst et al. (2015); Suffolk and Poortinga (2016); Suffolk (2016); Swim and Bloodhart (2013); Thøgersen (1999); Thøgersen and Noblet (2012); Thomas et al. (2016, 2019); Tiefenbeck et al. (2013); Tiefenbeck (2014); Tippet (2018); Truelove et al. (2014, 2016); Van der Werff et al. (2014a,b); Vasan (2018); Verfuerth and Gregory-Smith (2018); Vincent and Koessler (2018); Werfel (2017); Whitmarsh et al. (2018); Xanthopoulou and Papagiannidis (2012); Xie et al. (2017); Xu et al. (2018a,b); Zawadzki (2015); Zimmerman (2009).

		I've been lazy today, best not eat so much tonight	I've been lazy today, so, what the heck, let's have a big slice of cake

Other real world examples from environmental behaviour are whether a behavioural intervention to monetarily incentivise household waste separation has a significant effect not just on waste separation (behaviour 1), but also on green shopping, travelling, and support to environmental policies (behaviour 2), for instance (Xu, Zhang, and Ling, 2018); or whether an intervention to restrict irrigation has a significant impact not just on water conservation (behaviour 1), but also on recycling behaviour (behaviour 2), for example (Sintov, Geislar, and White, 2019).

The mechanisms thought to explain promoting or positive spillovers vary by discipline and theoretical framework. Psychological approaches have focussed particularly on two mechanisms: (a) self-perception, identity, or preference for consistency (behaviour 1 changes how one sees oneself and the desire to act consistently with that self-image leads to behaviour 2) and (b) self-efficacy, knowledge, or self-motivation/empowerment (satisfactorily undertaking behaviour 1 increases confidence and perceived efficacy of action, motivating change in behaviour 2; Nash et al., 2017). Permitting or negative spillovers have been typically explained in terms of moral licensing, whereby a virtuous initial behaviour licenses or 'permits' a second indulgent or morally-questionable behaviour, or by a contribution ethic whereby an initial behaviour justifies subsequent inaction (e.g., Thøgersen, 1999; Karmarkar and Bollinger, 2015). Rebound effects are a related phenomenon, studied more from an economic than psychological perspective, and describe increased energy consumption due to technical efficiency gains, thereby offsetting energy savings achieved (e.g., Sorrell & Dimitropoulos, 2008).

Evidence for spillover remains somewhat mixed, with some studies finding effects under certain conditions that are not replicated in other studies (Nash et al., 2017). Conceptually, spillover remains defined and explained in a variety of ways, and there remain considerable gaps in understanding (e.g., the role of social processes, such as norms, in spillover; Nash et al., 2017). Methodologically, there is also no coherent approach to researching spillover, which may in part explain the mixed and inconsistent empirical results, and critically highlights a need to improve the rigour and transparency of spillover research.

1.3 Overview of spillover research methods and measurement

A growing stream of the literature at the interface between economics and psychology is currently investigating 'behavioural spillovers' in (and across) different domains, including health, environmental, and pro-social behaviours. To date, there have been a variety of methods applied to studying spillover (see Table 1). These range from qualitative retrospective self-reports using biographical interviews (e.g., Nash et al., submitted) to controlled laboratory experiments with randomisation to condition (e.g., Van der Werff et al., 2014a,b). Each approach offers different strengths and weaknesses. For example, qualitative approaches are able to elucidate unexpected spillovers and additional insights (e.g., drivers, barriers, mechanisms) not anticipated or measured in quantitative approaches. On the other hand, quantitative approaches allow for more measurement standardisation and potentially for generalisation, as well as affording insights into factors shaping behaviour that individuals may be unable or unwilling to reflect on consciously through self-report.

Table 1. Overview of methods used to research behavioural spillover: examples from environmental behaviour

Methodological approach	**Data collection & analysis methods**	**Examples from environmental behaviour**	**Strengths**	**Weaknesses**
Qualitative	• Interviews or open-ended survey questions • Thematic, content, discourse (or similar) analysis • Self-reports or other (e.g., practitioner) accounts • Biographical (retrospective) or evaluative (during / immediately after intervention)	Austin et al., (2011); Nash et al., (submitted); Uzzell and Räthzel (2018); Boström et al. (2015)	• Expose unexpected spillovers • Shed light on spillover mechanisms, drivers and barriers	• Risk of presentational bias • Partial or selective recollection • No measurement standardisation
Quantitative (cross-sectional)	• Survey, card sort or secondary data analysis (e.g., retail data) • Cluster or factor analysis • Correlational analysis • Regression analysis	Thogersen, (1999); Barr et al., (2010); Whitmarsh and O'Neill, (2010); Austin et al., (2011); Gabe-Thomas et al., (2016).	• Quantify strength of relationships between measured behaviours • Measurement standardisation	• No causal relationships identified • Limited to expected spillovers
Quantitative (longitudinal)	• Surveys at 2+ timepoints • Repeated measures analysis or multi-level modelling • Correlational analysis • Regression analysis (including time series, panel data, and difference-in-difference models)	Noblet and Thøgersen (2012); Kaida and Kaida (2013); Poortinga et al., (2013); Thomas et al., (2016).	• Quantify strength of relationships between measured behaviours • Measurement standardisation	• No causal relationships identified • Limited to expected spillovers
Quantitative (experimental)	• Online, laboratory, or field experiments • Self-reported or observed behaviour • Randomisation to behavioural intervention • Analysis of variance • Regression analysis	Van der Werff et al., (2014a,b); Juhl et al., (2018).	• Causal relationships and mechanisms identified • Measurement standardisation	• Limited to expected spillovers
Mixed-methods	• Combination of qualitative and quantitative methods (e.g., experiment & interviews)	Verfuerth (in preparation); Lede (in preparation).	• As above	• As above

Measurement of spillover has been undertaken in a variety of ways that reflect the range of methods used. Qualitative approaches tend to rely on self-reported accounts of behaviour change; whereas quantitative approaches may use self-reports or observations of behaviour. A key weakness in the literature to date, has been a reliance on self-reported behaviour, which is

known to be only weakly correlated with actual behaviour (e.g., Kormos and Gifford, 2014). Furthermore, several studies claiming to find spillover have found change in behavioural intentions or attitudes following an initial behaviour change, which is not strictly spillover (Van der Werff et al., 2014). Few studies also conduct follow-up measurements, so the durability of any immediate spillover effects is unknown. There has also been a reliance on correlational or longitudinal designs which are unable to shed light on causal processes; and within longitudinal designs approaches differ in how to detect spillover (Capstick et al., submitted). Finally, there have also been few attempts to bring together quantitative and qualitative approaches, thus providing complementary insights and addressing respective weaknesses in approaches (Creswell, 2014). In the following section, we describe how spillover should be measured in experimental and non-experimental approaches that seeks to build on this literature and address limitations in the methods used to date.

2 Measuring spillover

We now turn from our observations of previous spillover research to a discussion of how we propose spillover research should ideally be conducted in order to reliably detect any spillover effects and expose mechanisms through which they may operate. Drawing on best practice in research design and reflecting principles of transparency and validity (e.g., Open Science Collaboration, 2015), we first discuss experimental studies, which elucidate causal mechanisms, and then non-experimental approaches, which afford other insights into spillover, as discussed above.

2.1 How to measure behavioural spillover: experimental studies

Rigorously designing and implementing randomised controlled experiments allows the researchers to obtain an unbiased estimate of the average treatment effect of a behavioural intervention (e.g. a 'nudge', a monetary or non-monetary incentive, a 'boost' or 'prime'). Because of sample selection bias, it is only by randomly assigning subjects to a treatment or to a control group that the researchers can identify the causal effect of a behavioural intervention on an observed outcome (Heckman, 1979; Burtless, 1995; Angrist and Pischke, 2009; List, 2011; Gerber and Green, 2012).

In practice, a variety of different randomised controlled experiments is available to researchers interested in testing behavioural spillovers. It is useful to refer here to the influential taxonomy of experiments in social sciences originally proposed by Harrison and List (2004): *conventional lab* experiments involve student subjects, abstract framing, a lab context, and a set of imposed rules; *artefactual field* experiments depart from conventional lab experiments in that they involve non-student samples; *framed field* experiments add to artefactual field experiments a field context in the commodity, stakes, task or information; and, finally, *natural field* experiments depart from framed field experiments in that subjects undertake the tasks in their natural environment, and subjects do not know that they take part into an experiment. The main idea behind *natural field* experiments is that the mere act of observation and measurement necessarily alters what is being observed and measured. In key areas of interest for behavioural spillovers, such as health, the environment or pro-social behavior, for instance, there are potential *experimenter demand effects* (i.e. participants change behaviour due to cues about what represents 'appropriate' behaviour for the experimenter: Bardsley, 2005; Levitt and List, 2007a; 2007b; Zizzo, 2010); *Hawthorne effects* (i.e. simply knowing they are part of a study makes participants feel important and improves their effort and performance: Franke and Kaul,

1978; Adair, 1984; Jones, 1992; Levitt and List, 2011); and *John Henry effects* (i.e. participants who perceive that they are in the control group exert greater effort because they treat the experiment like a competitive contest and they want to overcome the disadvantage of being in the control group: Campbell and Stanley, 1963; Cook and Campbell, 1979).

Other, more recent, typologies of randomised controlled experiments are *online experiments* (Horton et al., 2011) conducted, for instance, using Amazon's *Mechanical Turk* (*MTurk*) (Paolacci et al., 2010, 2014; Horton et al., 2011); and *lab-field experiments* that consist of a first-stage intervention under controlled conditions (in the lab) linked to a naturalistic situation (in the field) where subjects are not aware that their behaviour is actually observed. Lab-field experiments have been used to look at the unintended spillover effects of behavioural interventions in health (Dolan and Galizzi, 2014; 2015; Dolan, Galizzi, and Navarro-Martinez, 2015), as well as at the spillover effects in terms of external validity of lab-based behavioural economics games of pro-social behaviour (Galizzi and Navarro-Martinez, 2018).

Investigating experimentally the occurrence of behavioural spillover requires a mixed, longitudinal experimental design combining elements of between- and within-subjects design. Participants in an experiment are randomly allocated by the researcher either to a control group, or to (at least) one behavioural intervention group. In the control group (C), subjects are observed while they engage in a first behaviour (behaviour 1) and then in a different, subsequent, behaviour (behaviour 2). Each of the two subsequent behaviours is operationally captured and reflected into (at least) one corresponding outcome variable: B1 and B2. In practice, the choice of behaviour 1 and behaviour 2, as well as the choice of the corresponding outcome variables B1 and B2, is often based on theoretical expectations, previous literature, or qualitative evidence. It is also based on other, more pragmatic, considerations related, for example, to the ease of observing some specific positive or negative spillovers in the lab or the field, and to the ethical and logistical acceptability of changing some behaviours in an experimental setting. In what follows, we illustrate the measurement of behavioural spillovers in the simplest possible case of one single behavioural intervention group, and one single outcome variable for both B1 and B2. The extension to more complex cases is straightforward.

In the treatment group (T), a behavioural intervention (e.g. a 'nudge', a monetary or non-monetary incentive, a 'boost' or 'prime') is introduced to directly target behaviour 1, thus affecting the outcome variable B1. The between-subjects design naturally allows the researcher to test the effects of the behavioural intervention on the targeted behaviour 1, by directly comparing B1 across the control and the treatment groups, that is, by comparing B1C versus B1T.

The between-subjects design, together with the longitudinal dimension of the experiment, also allows the researcher to check if the behavioural intervention has a ramification effect on the non-targeted behaviour 2, thus affecting the outcome variable B2. In particular, the outcome of behaviour 2 in the control group (B2C) serves as the baseline level for the extent to which behaviour 2 is affected by behaviour 1 in the absence of any behavioural intervention targeting behaviour 1 (B1C).

In contrast, the outcome of behaviour 2 in the treatment group (B2T) captures the extent to which behaviour 2 is affected by the 'perturbed' level of behaviour 1 as a consequence of the introduction of the behavioural intervention (B1T).

Therefore, by directly comparing B2T and B2C, the difference $\Delta B2 = B2T - B2C$ captures the positive or negative change in the outcome variable for behaviour 2 which is directly attributable to the change in the outcome variable for behaviour 1, $\Delta B1 = B1T - B1C$, which, in turn, is causally affected by the introduction of the behavioural intervention. That is, $\Delta B2 = B2T - B2C$ captures the 'knock on' behavioural spillover effect of the behavioural intervention targeting behaviour 1 on the non-targeted, subsequent behaviour 2.

Figure 2: Experimental design and variables to test behavioural spillovers

	Behaviour 1	**Behaviour 2**
Control group (C)	B1C	B2C
Treatment group (T)	B1T	B2T
Difference	ΔB1	ΔB2

In terms of sizes and statistical significance, such spillover effects may not be significantly different from zero ($\Delta B2 = 0$), may be significantly and positively different from zero (i.e. $\Delta B2 > 0$), or, finally, may be significantly and negatively different from zero (i.e. $\Delta B2 < 0$). If the two behaviours share one common underlying 'motive' (in the sense of Dolan and Galizzi, 2015, of some overarching goal or deep preference such as 'being healthy', 'being pro-environmental', or 'being pro-social') then the experimental findings may thus be interpreted as evidence of no behavioural spillovers ($\Delta B2 = 0$), evidence of originating 'promoting' or 'precipitating' behavioural spillover ($\Delta B2 > 0$) or, finally, evidence of 'permitting' or 'purging' behavioural spillover ($\Delta B2 < 0$).

Such an experimental design also allows the researchers to estimate not only the sign and the statistical significance of the behavioural spillover effects, but also their size. In particular, by comparing the relative changes in the outcome variables for behaviour 1 and 2 as effects of the introduction of the behavioural intervention, the ratio between the proportional change ($\Delta B2/B2C$) and the proportional change ($\Delta B1/B1C$) allows the researcher to estimate the 'elasticity' of the behavioural spillovers: in analogy with standard price elasticity concepts, the elasticity is defined as the percentage change in behaviour 2 per unitary percentage change in behaviour 1, that is $\varepsilon_{BS} = (\Delta B2/B2C)/(\Delta B1/B1C)$.

This, in turn, allows the researcher to conclude whether a behavioural intervention causes behavioural ramifications which are small or large compared to the directly targeted change in behaviour. In case of permitting or purging behavioural spillovers (i.e. $\Delta B1$ and $\Delta B2$ having opposite signs), and provided that B1 and B2 share the same metrics (or provided that they feed into the underlying motive in a way that the relative sizes of their changes $\Delta B1$ and $\Delta B2$ are conceptually comparable), this can provide further evidence on whether the permitting or purging spillovers are compensating each other completely or only partially (e.g. 'backfire' or 'rebound' effects).

Two further considerations are in order here. First, the above described definition and framework to measure behavioural spillovers in an experimental setting is sufficiently general and comprehensive to nest as a special case the situation where the behavioural intervention consists of behaviour 1 itself. For example, in the 'question-behaviour' and 'survey' promoting spillover effects discussed in Dolan and Galizzi (2015), the behavioural intervention consists of randomly assigning subjects to a brief survey or questionnaire eliciting past health, environmental, or purchasing behaviour (e.g. Fitzsimons and Shiv, 2001; Zwane et al., 2011; Van der Werff, Steg, and Keizer, 2014). In such a case, in fact, the behavioural intervention in

the treatment group merely consists of exposing subjects to behaviour 1 (e.g. a survey) before behaviour 2 takes place. In the control group, on the other hand subjects go through behaviour 2 without being previously exposed to behaviour 1. Also in this, simpler, special case, behavioural spillover is measured as $\Delta B2 = B2T - B2C$, but in this case the behavioural spillover captures the positive or negative change in the outcome variable for behaviour 2 which is directly attributable to the mere exposure of subjects to behaviour 1 in the treatment group (which, in this case, coincides with the behavioural intervention).

Second, the decision about the timeframe is crucial for the measurement of behavioural spillovers. Following subjects over longer timeframes implies, naturally, that it is more likely that spillover effects are effectively detected (Poortinga et al., 2013). Considering substantially long timeframe (ideally a few weeks or even months after the end of the intervention) is desirable in order to be able to assess the durability of spillover effects. Considering even longer timeframes (ideally over three or six months after the end of the intervention) is particularly important to be able to detect the formation of new habits sustained over time (Lally et al., 2010), rather than a behavioural change that is only transient. In any case, in order to favour transparency and replicability of experimental results, it is crucial that the researchers pre-specify in advance the timeframe over which subjects are followed up over time. The timeframe, in fact, is a key point of the checklist that we propose below.

2.2 How to measure behavioural spillover: non-experimental quantitative studies

An analogous strategy can be used in non-experimental settings along the line of the difference-in-difference empirical approach (e.g. Card, 1992, 1996; Card and Krueger, 1994, 2000; see more below). In particular, the researcher can exploit the variation occurring naturally in the field outside their control and can use some 'natural experiment' as an exogenous 'intervention' in order to identify the likely effect of such an exogenous change on the variables of interest, despite the fact that participants are not randomly assigned to a proper experimental intervention.

The exogenous variation occurring naturally in the field can be a change in policy, a natural 'shock' (e.g. a health shock, a natural disaster, a political shock, an economic shock), a life event (e.g. birth of a child, death of a relative, divorce, unemployment), a technological advance, a discontinuity in the availability or in the access of a resource or an infrastructure. The source of the exogenous variation can also be 'cognitive' or 'behavioural', such as an exogenous change in attention or awareness, provided that there are convincing reasons to argue that such a source of variation is exogenous (rather than endogenous) to the occurrence of behavioural spillovers.

In the standard difference-in-difference approach, two areas (e.g. two regions, two countries, two schools, two hospitals), are compared before and after the occurrence of a natural event (e.g. a policy, a shock) affecting one area (T) but not the other one (C). Typically, the change of the outcome of behaviour 1 before (t=0) and after (t=1) the natural event in the 'control' area $B1C_{t=1} - B1C_{t=0}$ is compared over time to the analogous change in the 'treatment' area $B1T_{t=1} - B1T_{t=0}$, in order to see whether the trends show any significant difference in differences across the two areas (i.e. if $B1T_{t=1} - B1T_{t=0}$, is statistically significantly different from $B1C_{t=1} - B1C_{t=0}$).

In principle, an analogous comparison can be made considering the outcome variable of behaviour 2 (B2, instead of B1), to see whether the natural event also has ramifications on a different, subsequent behaviour, far and beyond the initial change on behaviour 1. Therefore, the researcher can compare the change over time of the outcome variable for behaviour 2 before (t=0) and after (t=1) the natural event in the 'control' area $B2C_{t=1} - B2C_{t=0}$ to the analogous change in the outcome variable for behaviour 2 in the 'treatment' area $B2T_{t=1} - B2T_{t=0}$, in order to see whether the trends show any significant difference in differences across the two areas (i.e. whether $B2T_{t=1} - B2T_{t=0}$, is statistically significantly different from $B2C_{t=1} - B2C_{t=0}$). Analogous considerations to the ones described above can be made here concerning the sign, significance, and size of the behavioural spillovers in a non-experimental setting (e.g. Claes and Miliute-Plepiene, 2018).

As mentioned above, our framework is sufficiently general and comprehensive to nest, as a special case, the situation where the 'intervention' in an experimental setting, or the 'shock' or exogenous variation in a non-experimental setting, consists of behaviour 1 itself. In such a case, the difference-in-difference approach described above reduces to the comparison of the change in the outcome variable for behaviour 2 in the 'treatment' area that has been exposed to behaviour 1 ($B2T_{t=1} - B2T_{t=0}$) with the analogous change in the 'control' area which has not been exposed to behaviour 1 ($B2C_{t=1} - B2C_{t=0}$).

The empirical strategy described above has been illustrated having in mind our specific definition of behavioural spillover proposed in section 1.2, that is, the observable and causal effect that a change in one behaviour (behaviour 1) has on a different, subsequent behaviour (behaviour 2). Nonetheless, a corresponding strategy can be adapted to some of the instances encompassed by the broader definition of spillover reported at the beginning of section 1.2, that is the impact that an intervention in a given domain (e.g. health, the environment), group, or location, has on a different domain, group or location. In principle, two locations (e.g. two countries), can be compared before and after the occurrence of a natural event (e.g. a natural phenomenon, an intervention) affecting one domain (e.g. the environment) in one area (T) but not in the other one (C). The researcher can compare not only the change over time of the outcome variable for the domain directly involved in the phenomenon or originally targeted by the intervention (e.g. the environment), but also the change over time of the outcome variable for a different domain (e.g. health). Considering the knock-on effects of the phenomenon or intervention on different groups or regions is also possible in principle, although in practice the empirical analysis would need to account for other underlying intra-groups or intra-regional differences between the 'control' and the 'treatment' areas.

2.3 How to study behavioural spillover: qualitative and mixed-methods studies

A different, but potentially complementary, approach to studying spillover involves using qualitative methods, such as interviews analysed thematically (e.g., Nash et al., submitted; Uzzell and Räthzel, 2018; Boström et al., 2015; Dittmer and Blazejewski, 2016; Thomas et al., 2019). As noted, such approaches have the advantage over quantitative approaches of exposing unexpected spillovers, as well as the shedding light on the drivers, barriers and mechanisms of spillover, and on participants' experience and meanings associated with spillover. For example, Uzzell and Räthzel (2018) used life history interviews to examine how equivalent practices (as well as identities and meanings) develop over time and may be transferred between work and home; using diachronic and synchronic analyses allowed them to identify drivers and barriers to consistency of actions across time, as well as across contexts. Verfuerth et al. (2018) used depth interviews to explore the impacts of a workplace meat reduction intervention, and found

unanticipated spillover across behaviours (e.g., to avoiding food waste) and contexts (to home); while Schütte and Gregory-Smith's (2015) semi-structured interviews exposed cognitive and emotional barriers to pro-environmental spillover between home and holiday.

As such, qualitative methods provide valuable insight in their own right into spillover phenomena, but can also be combined with quantitative approaches in mixed-methods designs to address quantitative limitations (Verfuerth and Gregory-Smith, 2018). Various approaches can be used to ensure the quality of qualitative data, such as member validation (i.e., asking participants to check researcher interpretations), inter-rater reliability of coded data (i.e., using multiple coders and resolving any disagreement in interpretation), and reflexivity (i.e., fully documenting the processes used to collect data and the role and background of the researcher; Breakwell et al., 2012). Others have noted that the diversity of qualitative methods requires a range of criteria for assessing quality and validity (Reicher, 2000); but most agree at least that transparency and consistency are key (Braun and Clark, 2006). The importance of being systematic is therefore a criterion of quality shared by both quantitative and qualitative methods.

A growing literature advocates the use of mixed-methods approaches in order to triangulate and provide complementary insights. Despite associations of qualitative and quantitative methods with divergent epistemological and ontological paradigms (Blaikie, 1991), this should not imply that qualitative and quantitative methods are essentially incommensurate (Bryman, 1988). Rather, the distinction between particular qualitative and quantitative methods can be understood as primarily technical, and not necessarily philosophical. Qualitative and quantitative methods offer different insights into spillover and each is better suited to answering different types of research question (e.g., What are the range of effects of an intervention? How is the development of identity and practices experienced over time and contexts? What causes and mediates spillover?). Thus, the rationale for combining methods stems from "the basic and plausible assertion that life is multifaceted and is best approached by the use of techniques that have a specialized relevance" (Fielding and Fielding, 1986, p.34). Furthermore, using multiple methods allows interesting lines of inquiry exposed through one method to be explored further through another (Whitmarsh, 2009). At the same time, however, it is not assumed that aggregating data sources can provide a complete or 'true' picture of the social world (Silverman, 2001). Indeed, "the differences between types of data can be as illuminating as their points of coherence" (Fielding & Fielding, 1986, p.31), for example leading to a re-examination of conceptual frameworks or assumptions (Tashakkori and Teddlie, 2003). The distinct challenges of researching spillover imply both qualitative and quantitative approaches are warranted to address different facets of the problem.

Mixed-methods designs may be sequential or concurrent, or both (Creswell, 2014). In the case of spillover studies, a mixed methods design might start with an initial qualitative and/or correlational phase to identify clusters of co-occurring behaviours which may indicate spillover, for which candidate behaviours (B1, B2, etc.) and the causal pathways connecting them can be examined in a subsequent experimental design, as outlined above. In addition, qualitative methods can be used alongside quantitative behavioural measures within the intervention phase to explore the experience, perceptions, and subjective wellbeing implications of the intervention, and to expose potentially unexpected spillover effects, as well as possible drivers, barriers, mechanisms, and mediating/moderating factors for any spillover. This might take the form of interviews with a sub-sample of experimental participants, or one or more open-ended questions in a post-intervention survey. Where spillover is detected through quantitative experimental methods, qualitative data may help explain why this effect

has occurred, and how this has been subjectively perceived and experienced. In the event that spillover is not detected via the experimental methods outlined above, qualitative methods may explain why not, or they may expose other, unquantified spillover effects. Qualitative, quantitative, and experimental methods should thus be seen as complementary, rather than substitute, empirical methods to explore and assess behavioural spillovers. So far, there exist few mixed-methods studies of spillover, but those that have been undertaken appear to demonstrate that a mixed methodology can elucidate multiple aspects of spillover processes and experiences (Barr et al., 2010; Thomas et al., 2019; Verfuerth et al., 2018).

3 A practical checklist

Exploring and detecting behavioural spillovers is a research and policy task which should be undertaken using a systematic and transparent approach, in the same spirit of, and closely in line with, the recent best practices favouring and advocating systematisation and transparency in psychological and behavioural sciences (Ioannidis, 2005; Higgins and Green, 2011; Simmons, Nelson, and Simonsohn, 2011; Miguel et al., 2014; Simonsohn, Nelson, and Simmons, 2014; Open Science Collaboration, 2015; Munafò et al., 2017). In the previous section, we outlined how this might be achieved using different research designs.

Abstracting from these exemplar designs, here we propose a checklist of points which should be explicitly stated and addressed by the researcher prior to undertaking of experimental and empirical analysis. The 20-item checklist is in line with, and in the same spirit of, other checklists designed to systematically assess the methodological quality of prospective studies, for example by the Cochrane Collaboration (Higgins and Green, 2011). The checklist is also in line with, and in the same spirit of, other more general checklists guiding researchers through pre-registration of studies and pre-analysis plans (e.g., the Open Science Framework: https://osf.io/). Once filled in, the checklist for a prospective study should be deposited in a dedicated website which is going to be launched with the publication of this special issue, and which will be available at: www.behaviouralspillovers.net. The website will also include a data template where data from deposited studies could be shared, collated, and combined in order to conduct collaborative systematic reviews and meta-analyses of the literature.

The twenty questions of the checklist are below. In what follows we briefly illustrate each question with a real case study, the recent study by Xu, Zhang, and Ling (2018) on household waste separation:

1. What are the setting and population of interest?
    - Four geographically adjacent communities in the Yuhang District of Hangzhou, Zhejiang Province, China.

2. Is this an experimental or a non-experimental study?
    - An experimental study (a framed field experiment).

3. If this is a non-experimental quantitative study, what is the empirical identification strategy (e.g. difference-in-difference)?
    - N/A.

4. If this is a quantitative study, what is the control group?

- The control group were participants in each community who were not exposed to any formal promotion of waste separation.

5. How have the behaviours been selected (e.g., existing literature, qualitative evidence)?
    - Based on previous findings and on the literature.

6. What is the targeted behaviour 1?
    - Sorting daily garbage and bringing it to waste collection sites.

7. What are the outcome variables for behaviour 1 (i.e., how will you measure behaviour 1)? (Please list them and briefly describe each outcome variable, indicating whether this is directly observed or self-reported behaviour.)
    - Difference in self-reported household waste collection before and after the interventions.

8. How many intervention groups there are?
    - Originally there were three intervention groups, but one condition ('mixed condition') was then excluded (see footnote 1 in page 28).

9. What are the behavioural interventions targeting behaviour 1? (Please list them and briefly describe each of them.)
    - In the Environmental Appeal (EA) condition participants were given three monthly 30-minutes presentations where they were informed about the environmental benefits of waste separation. In the Monetary Incentive (MI) condition participants were given three monthly 30-minutes presentations where they were informed that they could earn 'green scores' from a recycling firm if they sorted their daily garbage and brought it to waste collection sites. In the 'mixed condition' participants were given three monthly 30-minutes presentations where they were informed of both EA and MI (this condition was later excluded from the analysis).

10. What is the non-targeted behaviour 2?
    - A set of 25 self-reported environmental behaviours or self-reported willingness to engage in environmental behaviours, including both 'private-sphere' behaviours (e.g. green shopping, travelling) and 'public-sphere' behaviours (e.g. support to environmental policies, environmental citizenship actions).

11. What are the outcome variables for behaviour 2 (i.e., how will you measure behaviour 2)? (Please list them and briefly describe each outcome variable, indicating whether this is directly observed or self-reported behaviour.). If there are multiple outcome variables for behaviour 2, does the study correct for multiple hypotheses testing? (Please describe which correction is used).
    - All the outcome variables for the 25 environmental behaviours or willingness to engage in environmental behaviours are self-reported, and are collected by a monthly survey. There is no explicit correction for multiple hypotheses testing.

12. What is the expected underlying motive linking behaviour 1 and behaviour 2?
    - Pro-environmental identity (page 28).

13. What are the expected mechanisms moderating and/or mediating the changes in the outcome variables for behaviour 2?
    - The expected mechanisms are both promoting/positive behavioural spillovers such as the activation of a stronger pro-environmental identity, and permitting/negative behavioural spillovers such as moral licensing (page 28). Pro-environmental identity and environmental concern are expected to mediate promoting/positive spillovers. Relief of guilt is expected to mediate permitting/negative spillovers.

14. What is the expected time frame during which behavioural spillovers will be tested, and during which the durability of spillover and habit formation will be assessed?
    - The expected time frame is not explicitly mentioned, but participants are followed up for three months.

15. What is the expected participant attrition between behaviour 1 and behaviour 2?
    - There is no explicit discussion of expected attrition. However, attrition was not only high, but it was asymmetric across different conditions. At the end of the experiment (three months after), only 195 out of the 400 participants originally recruited remained in the study: 80 (out of 100) in the EA group, 36 (out of 100) in the MI group, and 79 (out of 100) in the control group (all the 100 participants in the mixed condition group were excluded).

16. What is the expected direction of the changes in the outcome variables for behaviours 1 and 2 between the intervention groups and the control group (i.e., are positive or negative spillovers expected)?
    - Both promoting/positive and permitting/negative spillovers were expected (page 28).

17. What are the expected sizes and standard errors of the changes in the outcome variables for behaviours 1 and 2 between the intervention groups and the control group?
    - There is no explicit discussion of the expected effect size or standard errors of the changes in the outcome variables for behaviours 1 and 2.

18. What is the minimum expected sample size to test and detect the occurrence of behavioural spillover?
    - The study recruits n=100 participants in each of the four groups, but there is no explicit justification of the minimum expected sample size to test and detect the occurrence of behavioural spillovers.

19. If collecting qualitative data, how will the quality of this data be ensured and assessed (e.g., reflexivity, consistency)?
    - A number of psychological constructs were collected (including four items to measure personal identification with environmental protection; three items to measure personal concern for the environment, ecology, and the earth; three items to measure feelings of disappointment, guilt, and regret for past environmentally unfriendly behaviours) and used in exploratory factor analysis, but no further qualitative data was collected.

20. If using mixed-methods approaches, how will insights from different methods be combined?
    - N/A.

4 Conclusion

We have critically reviewed the main methods to measure behavioural spillovers to date, and discussed their methodological strengths and weaknesses. We have proposed a consensus mixed-method approach which uses a longitudinal between-subject design together with qualitative self-reports: participants are randomly assigned to a treatment group where a behavioural intervention takes place to target behaviour 1, or to a control group where behaviour 1 takes place absent any behavioural intervention. A behavioural spillover is empirically identified as the effect of the behavioural intervention in the treatment group on a subsequent, not targeted, behaviour 2, compared to the corresponding change in behaviour 2 in the control group.

In the spirit of the pre-analysis plan, we have also proposed a systematic checklist to guide researchers and policy-makers through the main stages and features of the study design in order to rigorously test and identify behavioural spillovers, and to ensure transparency, reproducibility, and meta-analysis of studies.

While ours is arguably the first methodological note on how to measure behavioural spillovers, it has of course limitations. The main limitation is that our experimental and empirical identification strategy relies on our specific definition of behavioural spillover – i.e. the observable and causal effect that a change in one behaviour (behaviour 1) has on a different, subsequent behaviour (behaviour 2). As mentioned in section 1.2, broader definitions of spillover exist that can encompass attitudinal change, learning, interpersonal influences, and other disparate processes. While we have suggested here that a similar approach to ours (i.e., longitudinal mixed-methodology) might apply in these cases, there may be also be methodological considerations specific to each type of spillover that warrants its own methodological checklist. Even applying our more specific definition of behavioural spillover, it would be possible to define alternative methodological checklists that, for example, apply solely quantitative or qualitative methods (cf. Uzzell and Räthzel, 2018). However, as we have argued, we believe there is benefit in combining methods as they can offer different insights or address different research questions relating to spillover.

We would like to conclude by briefly mentioning a few other directions where we envisage promising methodological developments in the years to come. First, the current technological landscape naturally lends itself to a systematic measurement of behavioural spillovers in a variety of research and policy domains. Today an unprecedented richness of longitudinal data are routinely collected at an individual level in terms of online surveys, apps, smart phones, internet of things (IoT) and mobile devices, smart cards and scan data, electronic administrative records, biomarkers, and other longitudinal panels. This is creating, for the first time in history, an immense potential for following up individuals across different contexts and domains, and over time, for months, years, and even decades. This new technological landscape is also creating previously unexplored opportunities for 'behavioural data linking', that is, for the linkage of behavioural experiments with other sources of longitudinal data (Galizzi, 2017; Galizzi, Harrison and Miraldo, 2017; Galizzi and Wiesen, 2018; Krpan, Galizzi and Dolan, 2019). On the one hand, the scope for systematically testing the occurrence of behavioural

spillovers using rigorous empirical and experimental methods is therefore enormous. On the other hand, the endless wealth of research hypotheses, outcome variables, and data points makes even more important for researchers to embrace the best practices discussed above in order to ensure transparency, openness, and reproducibility of science.

Second, a promising methodological line of research about behavioural spillover concerns the rigorous investigation of the factors mediating and moderating the occurrence of behavioural spillover, for example in terms of accessibility (Sintov, Geislar and White, 2019). Further work in this direction is likely to develop also thanks to the triangulation of different sources of data enabled by the above described shift in the technological landscape.

All these future developments reinstate the importance of developing a collective discussion about clear and transparent methodological guidelines to measure behavioural spillovers. We hope that with the present article we have contributed to at least start such a discussion. The time is ripe to foster a collaborative endeavour to systematically test behavioural spillovers across all research and policy domains, contexts, and settings.

5 References

Adair, G. (1984). The Hawthorne effect: a reconsideration of the methodological artefact. *Journal of Applied Psychology*, 69, 334-345.

Alpizar, F., Norden, A., Pfaff, A., & Robalino, J. (2013a). Effects of exclusion from a conservation policy: negative behavioral spillovers from targeted incentives. Duke Environmental and Energy Economics Working Paper EE 13-06.

Alpizar, F., Norden, A., Pfaff, A., & Robalino, J. (2013b). Behavioral spillovers from targeted incentives: losses from excluded individuals can counter gains from those selected. Duke Environmental and Energy Economics Working Paper EE 13-07.

Angelovski, A., Di Cagno, D., Guth, W., Marazzi, F., & Panaccione, L. (2018). Behavioral spillovers in local public good provision: an experimental study. *Journal of Economic Psychology*, 67, 116-134.

Angrist, J.D., Pischke, J.S. (2009). *Mostly harmless econometrics: an empiricist's companion*. Princeton: Princeton University Press.

Austin A., Cox J., Barnett J., Thomas C. (2011). *Exploring catalyst behaviours: full Report: a report to the Department for Environment, Food and Rural Affairs*. Brook Lyndhurst.

Baca-Motes, K., Brown, A., Gneezy, A., Keenan, E.A., & Nelson, L.D. (2013). Commitment and behavior change: evidence from the field. *Journal of Consumer Research*, 39, 1070-1084.

Banerjee, R. (2016). Corruption, norm violation and decay in social capital. *Journal of Public Economics*, 137, 14-27.

Bardsley, N. (2005). Experimental economics and the artificiality of alteration. *Journal of Economic Methodology*. 12(2), 239-251.

Barr S., Shaw G., Coles T., & Prillwitz J. (2010). 'A holiday is a holiday': practicing sustainability, home and away. *Journal of Transport Geography*. 18(3):474-81.

Beaman, A. L.; Cole, C. M.; Preston, M.; Klentz, B.; Steblay, N. M. (1983). Fifteen years of foot-in-the door research: A meta-analysis. *Personality and Social Psychology Bulletin, 9* (2), 181–196.

Bech-Larsen, T., & Kazbare, L. (2014). Spillover of diet changes on intentions to approach healthy food and avoid unhealthy food. *Health Education*, 114(5), 367-377.

Bednar, J., Chen, Y., Liu, T. X., & Page, S. (2012). Behavioral spillover and cognitive load in multiple games: Experimental results. *Games and Economic Behavior*, 74, 12–31.

Bednar, J., Joones-Rooy, A., & Page, S.E. (2015). Choosing a future based on the past: institutions, behaviour, and path dependence. *European Journal of Political Economy*, 40(B), 312-332.

Bednar, J., & Page, S.E. (2018). When order affects performance: culture, behavioral spillovers, and institutional path dependence. *American Political Science Review*, 112(1), 82-98.

Blaikie, N. W. H. (1991). A critique of the use of triangulation in social research. *Quality and Quantity, 25,* 115-136.

Blanken, I., de Ven, N. van, & Zeelenberg, M. (2015). A meta-analytic review of moral licensing. *Personality and Social Psychology Bulletin, 41*(4), 540–558.

Boström, M., Gilek, M., Hedenström, E. & Jönsson, A. M. (2015). How to achieve sustainable procurement for "peripheral" products with significant environmental impacts. *Sustainability: Science, Practice and Policy, 11* (1), 21-31.

Bratt, C. (1999). Comsumers' environmental behaviour: generalized, sector-based, or compensatory? *Environment and Behavior*, 31, 28-44.

Breakwell, G., Smith J.A, & Wright, D.B. (2012). *Research Methods in Psychology, 4th Edn*. London: Sage.

Braun, V. & Clarke, V. (2006) Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77-101.

Bryman, A. (1988). *Quantity and Quality in Social Research*. London: Unwin Hyman.

Burtless, G. (1995). The case for randomised field trials in economic and policy research. *Journal of Economic Perspectives,* 9, 63-84.

Campbell, D.T., & Stanley, J.C. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.

Capstick S., Lorenzoni I., Corner A., & Whitmarsh L. (2014). Prospects for radical emissions reduction through behavior and lifestyle change. *Carbon management*. 5(4):429-45.

Capstick S.B., Whitmarsh L., & Nash N. (2019) Behavioural spillover in the UK and Brazil: Findings from two large-scale longitudinal surveys. *Journal of Environmental Psychology* (submitted).

Card, D. (1992). Using regional variation to measure the effect of the Federal minimum wage. *Industrial and Labor Relation Review*. 46, 22-37.

Card, D. (1996). The effect of unions on the structure of wages: a longitudinal analysis. *Econometrica.* 64, 957-979.

Card, D., & Krueger, A. (1994). Minimum wages and employment: a case study of the fast-food industry in New Jersey and Pennsylvania. *American Economic Review*. 84, 772-784.

Card, D., & Krueger, A. (2000). Minimum wages and employment: a case study of the fast-food industry in New Jersey and Pennsylvania: Reply. *American Economic Review*. 90, 1397-1420.

Carpenter, C.S., & Lawler, E.C. (2017). Direct and spillover effects of middle school vaccination requirements. NBER Working Paper No. 23107.

Carrico, A.R., Raimi, K.T., Truelove, H.B., & Eby, B. (2017). Putting your money where your mouth is: an experimental test of pro-environmental spillover from reducing meat consumption to monetary donations. *Environment and Behavior*, 50(7), 723-748.

Cason, T. M., & Gangadharan, L. (2013). Cooperation spillovers and price competition in experimental markets. *Economic Inquiry*, 51(3), 1715–1730.

Cason, T., Savikhin, A. C., & Sheremeta, R. M. (2012). Behavioral spillovers in coordination games. *European Economics Review*, 56, 233–245.

Chatelain, G., Hille, S.L., Sander, D., Patel, M., Hahnel, U.J.J., & Brosch,T. (2018). Feel good, stay green: positive affect promotes pro-environmental behaviors and mitigates compensatory "mental bookkeeping" effects. *Journal of Environmental Psychology*, 56, 3-11.

Claes, E., & Miliute-Plepiene, J. (2018). Behavioral spillovers from food-waste collection in Swedish municipalities. *Journal of Environmental Economics and Management*, 89, 168-186.

Cook, T.D., & Campbell, D.T. (1979). *Quasi-experimentation: Design and analysis for field settings*. Chicago: Rand McNally.

Cornelissen, G., Pandelaere, M., Warlop, L., & Dewitte, S. (2008). Positive cueing: promoting sustainable consumer behaviour by cueing common environmental behaviors as environmental. *International Journal of Research in Marketing*, 25, 46-55.

Creswell, J. (2014). *A Concise Introduction to Mixed Methods Research*. Sage Publications.

Crookes, R.D. (2017). Exploring the underlying processes and the long term effects of choice architecture. Columbia University Doctoral Thesis.

Dickinson, D.L., & Oxoby, R.J. (2011). Cognitive dissonance, pessimism, and behavioral spillover effects. *Journal of Economic Psychology*. 32(3), 295-306.

Dittmer, F. & Blazejewski, S. (2016). Sustainable at home – sustainable at work? The impact of pro-environmental life-work spillover effects on sustainable intra- or entrepreneurship. *Sustainable Entrepreneurship and Social Innovation*: 73-100.

Dolan, P., & Galizzi, M.M. (2014). Because I'm worth it. A lab-field experiment on spillover effects of incentives in health. LSE CEP Discussion Paper CEPDP1286.

Dolan, P., & Galizzi, M.M. (2015): Like ripples on a pond: behavioral spillovers and their consequences for research and policy. *Journal of Economic Psychology*, 47, 1-16.

Dolan, P., Galizzi, M.M., & Navarro-Martinez, D. (2015). Paying people to eat or not to eat? Carryover effects of monetary incentives on eating behavior. *Social Science and Medicine*. 133, 153-158.

Dutschke, E., Frondel, M., Schleich, J., & Vance, C. (2018). Moral licensing: another source of rebound? Ruhr Economic Papers No. 747.

Eby, B.D. (2016). Color me geeen: the influence of environmental identity labelling on spillovers in pro-environmental behaviors. University of Colorado Boulder Master Thesis.

Falk, A., Fischbacher, U., & Gachter, S. (2013). Living in two neighborhoods – social interaction effects in the laboratory. *Economic Inquiry*, 51(1), 563-578.

Fanghella, V., d'Adda, G., & Tavoni, M. (2019). On the use of nudges to affect spillovers in environmental behaviors. *Frontiers in Psychology* (in press).

Fenger, M.H.J. (2017). Essays on the dynamics of consumer behavior. Aarhus University Doctoral Thesis.

Fielding, N.G. & Fielding, J.L. (1986). *Linking Data.* Newbury Park, California: Sage Publications.

Fitzsimons, G.J., & Shiv, B. (2001). Non-conscious and contaminative effects of hypothetical questions on subsequent decision-making. *Journal of Consumer Research*, 334, 782-791.

Franke, R.H., & Kaul, J.D. (1978). The Hawthorne experiments: first statistical interpretation. *American Sociological Review, 43(5),* 623-643.

Freedman, J. L. & Fraser, S. C. (1966). Compliance without pressure: The foot-in-the-door technique. *Journal of Personality and Social Psychology, 4*(2), 195–202.

Gabe-Thomas E., Walker I., Verplanken B., & Shaddick G. (2016). Householders' Mental Models of Domestic Energy Consumption: Using a Sort-And-Cluster Method to Identify Shared Concepts of Appliance Similarity. *PloS One*. 11(7).

Galbiati, R., Henry, E., & Jacquemet, N. (2017). Spillovers, persistence, and learning: institutions and the dynamic of cooperation. CEPR Discussion Paper No. DP12128.

Galizzi, M.M. (2017). Behavioral aspects of policy formulation: experiments, behavioral insights, nudges. In: Howlett, M. & Mukherjee, I. (Eds.) *Handbook of Policy Formulation*. Handbooks of Research on Public Policy. Edward Elgar Publishing, Cheltenham, UK, 410-429.

Galizzi, M.M., Harrison, G.W., & Miraldo, M. (2017). Experimental methods and behavioural insights in health economics: estimating risk and time preferences in health. In: Baltagi, B. & Moscone, F. (Eds.) *Health Econometrics in Contributions to Economic Analysis*. Emerald Publishing, Bingley, UK.

Galizzi, M.M., & Navarro-Martinez, D. (2018). On the external validity of social preferences games: a systematic lab-field study. *Management Science* (in press).

Galizzi, M.M., & Wiesen, D. (2018). Behavioral experiments in health economics. In: Hamilton, J.H., Dixit, A., Edwards, S., & Judd, K., (Eds.) *Oxford Research Encyclopedia of Economics and Finance*. Oxford Research Encyclopedias. Oxford University Press, Oxford, UK.

Geller, E.S. (2001). From Ecological Behaviorism to Response Generalization: Where Should We Make Discriminations? *Journal of Organizational Behavior Management*, Vol. 21(4), 55-73.

Gerber, A.S., & Green, D.P. (2012). *Field experiments: Design, analysis, and interpretation*, New York: W.W. Norton & Company.

Ghesla, C., Grieder, M., & Schmitz, J. (2018). Nudge for good? Choice defaults and spillover effects. *Frontiers in Psychology* (in press).

Gholamzadehmir, M. (2016). The impact of moral action and moral values on moral judgement and moral behaviour. University of Sussex Doctoral Thesis.

Godoy, S., Morales, A.J., & Rodero, J. (2013). Competition lessens competition: an experimental investigation of simultaneous participation in a public good game and a lottery contest game with shared endowment. *Economics Letters*, 120(3), 419-423.

Goswami, I. & Urminsky, O. (2015). The nature and extent of post-reward crowding-out: the 'effort balancing' account. SSRN Working Paper 2733335.

Ha, S., & Kwon, S.Y. (2016). Spillover from past recycling to green apparel shopping behaviour: the role of environmental concerns and anticipated guilt. *Fashion and Textiles*, 3:16. https://doi.org/10.1186/s40691-016-0068-7.

Harrison, G.W., & List, J.A. (2004). Field experiments. *Journal of Economic Literature*, 42, 1009-1055.

Hecht, T.D., & Boies, K. (2009). Structure and correlates of spillover from nonwork to work: an examination of nonwork activities, well-being, and work outcomes. *Journal of Occupational Health Psychology*, 14(4), 414-426.

Heckman, J.J. (1979). Sample selection bias as a specification error. *Econometrica,* 47, 153-162.

Hedrick, V.E., Davy, B.M., You, W., Porter, K.J., Estabrooks, P.A., & Zoellner, J.M. (2017). Dietary quality changes in response to a sugar-sweetened beverage-reduction intervention: results from the Talking Heath randomized controlled clinical trial. *American Journal of Clinical Nutrition*, 105(4), 824-833.

Hertwich, E.G. (2005). Consumption and the rebound effect: an industrial ecology perspective. *Journal of Industrial Ecology*, 9(1-2), 85-98.

Higgins, J.P.T., & Green, S. (2011). Cochrane Handbook for Systematic Reviews of Interventions. *Cochrane Collaboration*: http://handbook-5-1.cochrane.org/front_page.htm.

Horton, J., Rand, D., & Zeckhauser, R. (2011). The online laboratory: conducting experiments in a real labor market. *Experimental Economics*, 14(3), 399-425.

Ioannidis, J.P. (2005). Why most published research findings are false. *PLoS Medicine*, 2, e124.

Jessoe, K., Lade, G.E., Loge, F., & Spang, E. (2017). Spillover from behavioral interventions: experimental evidence from water and energy use. E2e Working Paper 033.

Jones, S.R.G. (1992). Was there a Hawthorne effect? *American Journal of Sociology.* 98, 451-148.

Juhl, H.J., Fenger, M.H.J., & Thøgersen, J. (2017). Will the consistent organic food consumer step forward? An empirical analysis. *Journal of Consumer Research*, 44(3), 519-535.

Juvina, I., Saleem, M., Martin, J.M., Gonzalez, C., & Lebiere, C. (2013). Reciprocal trust mediates deep transfer of learning between games of strategic interaction. *Organizational Behavior and Human Decision Processes*, 120(2), 206-215.

Kaida N., & Kaida K. (2015). Spillover effect of congestion charging on pro-environmental behavior. *Environment, Development and Sustainability*. 7(3):409-21.

Karmarkar U.R., & Bollinger B. (2015). BYOB: How bringing your own shopping bags leads to treating yourself and the environment. *Journal of Marketing*. 79(4):1-5.

Karremans, J.C., Van Lange, P.A.M., & Holland, R.W. (2005). Forgiveness and its associations with prosocial thinking, feeling, and doing beyond the relationship with the offender. *Personality and Social Psychology Bulletin*, 31(10), 1315-1326.

Kesternich, M., Roemer, D., & Flues, F. (2017). The power of active choice: field experimental evidence on repeated contribution decisions to a carbon offsetting program. ZEW-Centre for European Economic Research Discussion Paper No. 16-091.

Klein, F. (2017). A drop in the ocean? Behavioral spillover effects and travel mode choice. Swedish University of Agricultural Sciences Uppsala Master Thesis.

Kormos, C., & Gifford, R. (2014). The validity of self-report measures of pro-environmental behavior: A meta-analytic review. *Journal of Environmental Psychology.* 40, 359–371.

Krieg, J., & Samek, A. (2017). When charities compete: a laboratory experiment with simultaneous public goods. *Journal of Behavioral and Experimental Economics*, 66, 40-57.

Krpan, D., Galizzi, M.M., & Dolan, P. (2019). Looking at behavioural spillovers at the mirror: the ubiquitous case for behavioural spillunders (submitted).

Lacasse, K. (2015). Addressing the "go green" debate: policies that encourage small green behaviors and their political spillover effects. *Journal of Science Policy and Governance*, 3(1), 1-33.

Lacasse, K. (2016). Don't be satisfied, identify! Strengthening positive spillover by connecting pro-environmental behaviors to an "environmentalist" label. *Journal of Environmental Psychology*, 48, 149-158.

Lacasse, K. (2017). Can't hurt, might help: examining the spillover effects from purposefully adopting a new pro-environmental behaviour. *Environment and Behavior*. https://doi.org/10.1177/0013916517748164.

Lanzini, P., & Thøgersen, J. (2014). Behavioral spillover in environmental domain: An intervention study. *Journal of Environmental Psychology*, 40, 381–390.

Lally, P., van Jaarsveld, C., Potts, H. & Wardle, J. (2010). How habits are formed: Modelling habit formation in the real world. *European Journal of Social Psychology.* 40 (6), 998-1009.

Lauren, N., Fielding, K.S., Smith, L., & Louis, W.R. (2016). You did, so you can and you will: self-efficacy as a mediator of spillover from easy to more difficult pro-environmental behavior. *Journal of Environmental Psychology*, 48, 191-199.

Lauren, N., Smith, L.D., Louis, W.R., & Dean, A.J. (2017). Promoting spillover: how past behaviors increase environmental intentions by cueing self-perceptions. *Environment & Behavior*, in press.

Lawler, E.C. (2018). Three essays in health economics: evidence from U.S. vaccination policy. Vanderbilt University Doctoral Thesis.

Levitt, S.D., & List, J.A. (2007a). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives*, 21(2), 153-174.

Levitt, S.D., & List, J.A. (2007b). Viewpoint: on the generalizability of lab behaviour to the field. *Canadian Journal of Economics*, 40(2), 347-370.

Levitt, S.D., & List, J.A. (2011). Was there really a Hawthorne effect at the Hawthorne Plant? An analysis of the original illumination experiments. *American Economic Journal: Applied Economics,* 3, 224-238.

List, J.A. (2011). Why economists should conduct field experiments and 14 tips for pulling one off. *Journal of Economic Perspectives*, 25(3), 3-16.

Littleford, C., Ryley, T.J., & Firth, S.K. (2014). Context, control, and the spillover of energy use behaviours between office and home settings. *Journal of Environmental Psychology*, 40, 157-166.

Liu, T.X., Bednar, J., Chen, Y., & Page, S. (2018). Directional behavioural spillover and cognitive load effects in multiple repeated games. *Experimental Economics* (in press).

Ludwig, T.D. & Geller, E.S. (1997) Assigned versus participative goal setting and response generalization: managing injury control among professional pizza deliverers. *Journal of Applied Psychology, 82,* 253.

Margetts, E.A., & Kashima, Y. (2016). Spillover between pro-environmental behaviors: the role of resources and perceived similarity. *Journal of Environmental Psychology*, 49, 30-42.

McCoy, D., & Lyons, S. (2017). Unintended outcomes of electricity smart-metering: trading-off consumption and investment behaviour. *Energy Efficiency*, 10(2), 299-318.

Miguel, E., Camerer, C.F., Casey, K., Cohen, J., Esterling, K.M., Gerber, A., Glennerster, R., Green, D.P., Humphreys, M., Imbens, G., Laitin, D., Madon, T., Nelson, L., Nosek, B.A., Petersen, M., Sedlmayr, R., Simmons, J.P., Simonsohn, U., & Van der Laan, M. (2014). Promoting transparency in social science research. *Science*, 343(6166), 30-31.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., & the PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the PRISMA Statement. *Annals of Internal Medicine*, 151(4), 264-269.

Mullen, E. & Monin B. (2016). Consistency versus licensing effects of past moral behavior. *Annual Review of Psychology, 67,* 363–385.

Munafò, M.R., Nosek, B., Bishop, D.V.M., Button, K.S., Chambers, C.D., Percie Du Sert, N., Simonsohn, U., Wagenmakers, E.J., Ware, J.J., & Ioannidis, J.P.A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, 1(1), 0021.

Nash, N., Whitmarsh, L., Capstick, S., Hargreaves, T., Poortinga, W., Thomas, G., Sautkina, E., & Xenias, D. (2017). Climate-relevant behavioural spillover and the potential contribution of social practice theory. *WIREs Climate Change.* 8:e481.

Nash, N., Capstick, S. & Whitmarsh, L. (2018). Behavioural spillover in context: Negotiating environmentally-responsible lifestyles in Brazil, China and Denmark. *Frontiers in Psychology*.

Nilsson, A., Bergquist M., & Schultz W.P. (2016). Spillover effects in environmental behaviors, across time and context: a review and research agenda. *Environmental Education Research*. 3:1-7.

Nolan, J.M. (2011). The cognitive ripple of social norms communications. *Group Processes & Intergroup Relations*, 14(5), 689-702.
Norden, A. (2013). Essays on behavioral economics and policies for provision of ecosystem services. Goteborg University Doctoral Thesis.
Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251).
Panos, M. (2018). Treat yourself or promote your health: a presentation and examination of the mechanisms behind health behavior spillover. University of Minnesota Doctoral Thesis
Paolacci, G., Chandler, J., & Ipeirotis, P.G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*. 5, 411-419.
Paolacci, G., Chandler, J. (2014). Inside the Turk: understanding Mechanical Turk as a participant pool. *Current Directions in Psychological Science*. 23(3), 184-188.
Peters, A.M., van der Werff, E., & Steg, L. (2018). Beyond purchasing: electric vehicle adoption motivation and consistent sustainable energy behaviour in the Netherlands. *Energy Research and Social Science*, 39, 234-247.
Polizzi di Sorrentino, E., Woelbert, E., & Sala, S. (2016). Consumers and their behavior: state of the art in behavioral science supporting use phase modeling in LCA and ecodesign. *International Journal of Life Cycle Assessment*, 21(2), 237-251.
Poortinga, W., Whitmarsh L., & Suffolk C. (2013). The introduction of a single-use carrier bag charge in Wales: Attitude change and behavioural spillover effects. *Journal of Environmental Psychology*. 36:240-7.
Poroli, A., & Huang, L. V. (2018). Spillover effects of a university crisis: A qualitative investigation using situational theory of problem solving. *Journalism & Mass Communication Quarterly, 95*(4), 1128–1149.
Reicher, S. (2000). Against methodolatry: some comments on Elliott, Fischer, and Rennie. *British Journal of Clinical Psychology, 39*, 1-6.
Rodriguez-Muñoz, A., Sanz-Vergel, A. I., Demerouti, E., & Bakker, A. B. (2014). Engaged at work and happy at home: a spillover-crossover model. *Journal of Happiness Studies,* 15(2), 271-283.
Santarius, T., & Soland, M. (2018). How technological efficiency improvements change consumer preferences: towards a psychological theory of rebound effects. *Ecological Economics*, 146, 414-424.
Savikhin, A. (2010). Essays on experimental and behavioral economics. Purdue University Doctoral Thesis.
Savikhin, A. C., & Sheremeta, R. M. (2013). Simultaneous decision-making in competitive and cooperative environments. *Economic Inquiry*, 51(2), 1311–1323.
Schatzki, T. (2010). *The Timespace of Human Activity.* Lexington Books**.**
Schmitz, J. (2018). Temporal dynamics of pro-social behaviour – an experimental analysis. *Experimental Economics* (in press).
Schütte, L., & Gregory-Smith, D. (2015). Neutralisation and mental accounting in ethical consumption: the case of sustainable holidays. *Sustainability*, 7, 1-14.
Seebauer, S. (2018). The psychology of rebound effects: explaining energy efficiency rebound behaviours with electric vehicles and building insulation in Austria. *Energy Research & Social Science*, 46, 311-320.
Sheremeta, R.M., Cason, T.N., & Samek. A. (2010). Behavioral spillovers in coordination games. SSRN Working Paper 1755609.
Shreedhar, G. (2018). Experiments in behavioural environmental economics, London School of Economics and Political Science Doctoral Thesis.

Shreedhar, G., & Mourato, S. (2018). Do biodiversity conservation videos cause pro-environmental spillover effects? Grantham Research Institute on Climate Change and the Environment Working Paper No. 302.

Silverman, D. (2001). *Interpreting Qualitative Data: methods for analysing talk, text and interaction. 2nd Edition*. London: Sage Publications.

Simmons, J.P., Nelson, L.D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22, 1359-1366.

Simonsohn, U., Nelson, L.D., & Simmons, J.P. (2014). P-curve and effect size: Correcting for publication bias using only significant results. *Perspectives on Psychological Science*, 9, 666-681.

Sintov, N., Geislar, S., & White, L.V. (2019). Cognitive accessibility as a new factor in pro-environmental spillover: results from a field study of household food waste management. *Environment and Behavior*, 51(1), 50-80.

Sorrell, S., Dimitropoulos J., & Sommerville M. (2009). Empirical estimates of the direct rebound effect: A review. *Energy policy*. 37(4):1356-71.

Spence, A., Leygue, C., Bedwell, B., & O'Malley, C. (2014). Engaging with energy reduction: does a climate change frame have the potential for achieving broader sustainable behaviour? *Journal of Environmental Psychology*, 38, 17-28.

Steinhorst, J., Klockner, C.A., & Matthies, E. (2015). Saving electricity – for the money or the environment? Risks of limiting pro-environmental spillover when using monetary framing. *Journal of Environmental Psychology*, 43, 125-135.

Steinhorst, J., & Matthies, E. (2016). Monetary or environmental appeals for saving electricity? Potentials for spillover on low carbon policy acceptability. *Energy Policy*, 93, 335-344.

Suffolk, C. (2016). Rebound and spillover effects: occupant behaviour after energy efficiency improvements are carried out. Cardiff University Doctoral Thesis.

Suffolk, C., & Poortinga, W. (2016). Behavioural changes after energy efficiency improvements in residential properties. In: Santarius T., Walnum, H., Aaal, C. (Eds.) Rethinking Climate and Energy Policies, 121-142.

Swim, J.K., & Bloodhart, B. (2013). Admonishment and praise: interpersonal mechanisms for promoting proenvironmental behaviour. *Ecopsychology*, 5(1), 20-31.

Tashakkori, A. & Teddlie, C. (2003). *Handbook of Mixed Methods in Social and Behavioral Research.* Thousand Oaks, California: Sage Publications.

Thøgersen J. (1999). Spillover processes in the development of a sustainable consumption pattern. *Journal of economic psychology*. 20(1):53-81.

Thøgersen J., & Noblet C. (2012). Does green consumerism increase the acceptance of wind power*? Energy Policy*. 51:854-62.

Thomas G.O., Poortinga W., & Sautkina E. (2016). The Welsh Single-Use Carrier Bag Charge and behavioural spillover. *Journal of Environmental Psychology*. 47:126-35.

Thomas, G.O, Sautkina, E., Poortinga, W., Wolstenholme, E. & Whitmarsh, L. (2019) The English plastic bag charge changed behaviour and increased support for other charges to reduce plastic waste. *Frontiers in Psychology,* in press.

Tiefenbeck, V., Staake, T., Roth, K., & Sachs, O. (2013). For better or for worse? Empirical evidence of moral licensing in a behavioral energy conservation campaign. *Energy Policy*, 57, 160–171.

Tiefenbeck, V. (2014). Behavioral interventions to reduce residential energy and water consumption: impact, mechanisms, and side effects. ETH Zurich Doctoral Thesis.

Tippet E. (2018). Behavioral spillover across health and environmental domains. Ohio State University Master Thesis.

Truelove H.B., Carrico A.R., Weber E.U., Raimi K.T., & Vandenbergh M.P. (2014). Positive and negative spillover of pro-environmental behavior: an integrative review and theoretical framework. *Global Environmental Change*. 29:127-38.

Truelove, H.B., Yeung, K.L., Carrico, A.R., Gillis, A.J., & Raimi, K.T. (2016). From plastic bottle recycling to policy support: an experimental test of pro-environmental spillover. *Journal of Environmental Psychology*, 46, 55-66.

Uzzell, D. & Räthzel, N. (2018). Border crossing and the logics of space: A case study in pro-environmental practices. *Frontiers in Psychology, in press.*

Van der Werff E., Steg L., & Keizer K. (2014a). I am what I am, by looking past the present the influence of biospheric values and past behavior on environmental self-identity. *Environment and Behavior*. 46(5):626-57.

Van der Werff, E., Steg, L., & Keizer, K. (2014b). Follow the signal: when past pro-environmental actions signal who you are. *Journal of Environmental Psychology*, 40, 273-282.

Vasan, A.A. (2018). Affect as a model of pro-environmental spillover. Vanderbilt University Master Thesis.

Verfuerth, C., Jones, C., Gregory-Smith, D. & Oates, C. (2018). Spillover effects through the lens of identity process theory - a qualitative approach. *International Association of People-Environment Studies*, July 9th-13th, Rome.

Verfuerth, C. & Gregory-Smith, D. (2018). Spillover of pro-environmental behaviour. In Wells, V.K., Gregory-Smith, D. & Manika, D. (Ed). *Research Handbook on Employee Pro-environmental Behaviour.* Edward Elgar: Cheltenham.

Vincent, M., & Koessler, A.K. (2018). Mapping moral pluralism in behavioural spillovers: a cross-disciplinary account of the multiple ways in which we engage in moral valuing. ZWB – Leibniz Information Centre for Economics Discussion Paper, Kiel, Hamburg.

Werfel, S.H. (2017). Household behaviour crowds out support for climate change policy when sufficient progress is perceived. *Nature Climate Change*, 7(7), 512.

Whitmarsh, L. (2009). What's in a name? Commonalities and differences in public understanding of "climate change" and "global warming". *Public Understanding of Science, 18,* 401–420.

Whitmarsh, L.E., Haggar, P., & Thomas, M. (2018). Waste reduction behaviors at home, at work, and on holiday: what influences behavioral consistency across contexts? *Frontiers in Psychology* (in press).

Whitmarsh L., & O'Neill, S. (2010). Green identity, green living? The role of pro-environmental self-identity in determining consistency across diverse pro-environmental behaviours. *Journal of Environmental Psychology*. 30(3):305-14.

Xanthopoulou, D., & Papagiannidis S. (2012). Play online, work better? Examining the spillover of active learning and transformational leadership. *Technological Forecasting and Social Change*, 79(7), 1328-1339.

Xie, X., Lu, Y., & Gou, Z. (2017). Green building pro-environment behaviors: are green users also green buyers? *Sustainability*, 9(10), 1703.

Xu, L., Zhang, X., & Ling, M. (2018a). Pro-environmental spillover under environmental appeals and monetary incentives: evidence from an intervention study on household waste separation. *Journal of Environmental Psychology*, 60, 27-33.

Xu, L., Zhang, X., & Ling, M. (2018b). Spillover effects of household waste separation policy on electricity consumption: evidence from Hangzhou, China. *Resources, Conservation and Recycling*, 129, 219-231.

Zizzo, D.J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, 13, 75-98.

Zawadzki, S.J. (2015). The interdependent nature of environmental behaviors: testing a conceptual framework for behavioral spillover. Pennsylvania State University Master Thesis.

Zimmerman, K.L. (2009). Operationalizing the antecedents of work-family positive spillover: a longitudinal study. Portland State University Doctoral Thesis.

Zwane, A.P., Zinman, J., Van Dusen, E., Pariente, W., Null, C., Miguel, E., Kremer, M., Karlan, D.S., Hornbeck, R., Gine, X. Duflo, E., Devoto, F., Crepon, B., & Banerjee, A. (2011). Being surveyed can change later behaviour and related parameter estimates. *Proceedings of the National Academy of Science United States of America*, 108(5), 1821-1826.

Appendix A. Methodology of systematic review of the literature.

In conducting and reporting our systematic review of the literature, we followed as closely as possible the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist and guidelines (Moher et al., 2009), as explained below.

A.1. Search strategy and key terms

Google Scholar was searched in December 2018 using the following combinations of exact phrases in the advanced search settings:

1) "behavioral spillover" (field TX all text) OR
2) "behavioral spillover" (field TX all text).

A.2. Selection and exclusion criteria

The authors reviewed and assessed all the references systematically, following a two-stage strategy. In the first stage, the inclusion criteria were applied to the title, the keywords, and the abstract; in the second stage, the criteria were applied to the abstract and the full text. All the papers were independently assessed for inclusion by each of the authors. Differences in opinions between the authors were solved through discussion.
The two stages worked as follows. In the first stage, a study was included only if it satisfied the following three criteria:

1) The study was available (no broken link).
2) The study was written in English.
3) The study presented new scientific material, in terms of: new empirical evidence or original experimental analysis of behavioural spillover; new theoretical definitions or conceptual frameworks for behavioural spillovers; systematic reviews or meta-analyses of existing studies on behavioural spillovers. This criterion excluded non-systematic reviews, commentaries, editorials, letters, or similar items.

Each article was sequentially evaluated against the three criteria, starting with criterion one and ending on criterion three. Whenever a criterion was not met, the article was excluded.
In the second stage, the abstract and the full text of the studies shortlisted in the first stage were screened, evaluated, and finally included according to two further criteria:

4) The study considered human behaviour.
5) The study used a definition of behavioural spillover substantially in line with our operational definition in section 1.2, that is, the observable and causal effect that a change in one behaviour (behaviour 1) has on a different, subsequent behaviour (behaviour 2).

We included both published and unpublished studies, for example studies in working paper or in dissertation form. If both published and unpublished versions of the study were available, we considered the published version. If different dates of the unpublished versions were available, we considered the most recent one.
To ensure that the set of studies retrieved was exhaustive and comprehensive, for each included study we also back-tracked and screened all the references cited in the article, applying the same inclusion criteria explained above.

A.3. Search results

The initial Google Scholar search resulted in a total number of n=529 entries on 17th December 2018 (n=305 for "behavioural spillovers" and n=224 for "behavioural spillovers". After n=51 duplicates were removed, the resulting number of studies was n=478. We then excluded the papers that were not accessible (n=16), were not written in English (n=11), or did not present new scientific material (n=97). A total of n=354 studies met all three criteria in this first stage of our selection strategy.

The abstract and the full text of the n=354 studies shortlisted were then screened and evaluated. We then excluded the studies that did not focus on human behaviour (n=17), and the studies whose definitions of behavioural spillovers was substantially different from our operational definition - or which did not define behavioural spillovers at all (n=240). A total of n=97 studies matched all the inclusion criteria in this second stage.

Back-tracking, screening, and evaluating the references cited in these n=97 articles against the same inclusion criteria retrieved further n=9 studies. So, at the end of the whole process, the systematic review resulted in a total of n=106 selected studies.

Of the n=106 selected studies, n=12 are Doctoral theses, n=5 are Master theses, and n=12 are still unpublished works, all which shows the growing interest on behavioural spillovers.

The selection process and the number of papers excluded and included in each stage are summarised in the PRISMA flow chart in Figure A1.

[Insert Figure A1 here]

Figure A1: PRISMA flow diagram of systematic review